A Publication of the American Psychological Association

Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

ARTICLES

The Restricting Effects of Awareness: A Paradox and an Explanation ... 163
Donald P. Spence and Bert Holland

Some Effects of Rejection upon Attraction to a Group 175
J. Diedrich Snoek

Similarity in Performance as Influenced by Interaction, Self-Esteem, and Birth Order 183
Ezra Stotland and Nicholas B. Cottrell

Dimensions of Subjective Adjustment 192
Joseph Veroff, Sheila Feld, and Gerald Gurin

Differences and Disagreement as Factors in Creative Group Problem Solving 206
L. Richard Hoffman, Ernest Harburg, and Norman R. F. Maier

Perceptual Recognition as a Function of Anxiety, Motivation, and the Testing Situation 215
E. D. Longenecker

Significant Factors in Hypnotic Behavior 222
Theodore Xenophon Barber and Louis Benjamin Glass

CRITIQUE AND NOTES

The Relationship between Response Style and Personality Variables: I. The Measurement of Response Acquiescence 229
Richard K. McGee

Expressive Movements, Warmth, and Verbal Reinforcement 234
Michael M. Reece and Robert N. Whitman

Overevaluation of Own Group's Product in Intergroup Competition 237
Robert R. Blake and Jane Srygley Mouton

Time Orientation in Male and Female Juvenile Delinquents 239
Anthony Davids, Catherine Kidder, and Melvyn Reich

Volume 64 MARCH 1962 Number 3

Volume 64 March 1962 Number 3

Journal of Abnormal and Social Psychology

Editor: DANIEL KATZ, *University of Michigan*

Associate Editor: ROBERT B. ZAJONC, *University of Michigan*

CONSULTING EDITORS

JOSEPH B. ADELSON
University of Michigan

GORDON W. ALLPORT
Harvard University

NORMAN H. ANDERSON
University of California, Los Angeles

LEONARD BERKOWITZ
University of Wisconsin

JACK W. BREHM
Duke University

URIE BRONFENBRENNER
Cornell University

DONALD R. BROWN
Bryn Mawr College

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
System Development Corporation

DWIGHT W. CHAPMAN, JR.
Vassar College

MORTON DEUTSCH
Bell Telephone Laboratories

SEYMOUR FESHBACH
University of Pennsylvania

DONALD W. FISKE
University of Chicago

NATHANIEL L. GAGE
University of Illinois

NICHOLAS HOBBS
Peabody College for Teachers

MURRAY HORWITZ
New York University

IRVING L. JANIS
Yale University

HAROLD H. KELLEY
University of California, Los Angeles

HERBERT C. KELMAN
Harvard University

DANIEL J. LEVINSON
Harvard Medical School

DAVID C. MCCLELLAND
Harvard University

ROBERT B. MACLEOD
Cornell University

HAROLD L. RAUSH
National Institute of Mental Health

NEVITT SANFORD
Stanford University

IRVING SARNOFF
Western Reserve University

WILLIAM A. SCOTT
University of Colorado

JOSEPH VEROFF
University of Michigan

JUSTIN L. WEISS
Massachusetts Mental Health Center

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued monthly, two volumes per year. The subscription price per volume is $10.00, or $20.00 annually (foreign, $20.50). Single issues are $2.00.

Original manuscripts should be sent to the Editor:

DANIEL KATZ
University of Michigan
Box 1248
Ann Arbor, Michigan

Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts may be obtained from the Editor or from the APA Central Office.

HELEN ORR
Acting Production Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Sts., Lancaster, Pa.
and 1333 Sixteenth Street N.W., Washington 6, D. C.

Second-class postage paid at Lancaster, Pa., and at additional mailing offices

Copyright © 1962 by the American Psychological Association, Inc.

Journal of Abnormal and Social Psychology
1962, Vol. 64, No. 3, 163–174

THE RESTRICTING EFFECTS OF AWARENESS:

A PARADOX AND AN EXPLANATION [1]

DONALD P. SPENCE AND BERT HOLLAND

Research Center for Mental Health, New York University

An increasing body of data in recent years has shown that a very brief or faint perceptual stimulus may have an effect on behavior, but the nature of this effect and ways to account for it are still under discussion. The effect is generally slight (as in studies by Klein, Spence, Holt, & Gourevitch, 1958; Paul & Fisher, 1959; Smith, Spence, & Klein, 1959) and in some cases open to the suspicion that partial awareness of the stimulus could account for the response. Such critics as Eriksen have argued that some aspects of the stimulus were consciously perceived but unverbalized; since we depend on a verbal response ("I didn't see anything") to measure awareness, we are misled into thinking that no conscious cues are being received. Because the insensitive verbal indicator is used to determine threshold, it leads to the false conclusion that a stimulus is subliminal. In reality, says Eriksen (1960), the subject is receiving conscious information which he cannot report. Because the cues are fragmentary, the response is usually inexact, in proportion to the indistinct quality of the stimulus.

The study by Smith et al. (1959) can be used to illustrate this argument. Two words, HAPPY and ANGRY, were briefly exposed and immediately followed by a consciously seen, ambiguous face. Descriptions of the face appeared to be influenced by the unseen verbal stimulus but the effects were indirect, the descriptions rarely containing the words HAPPY or ANGRY. Since a later check showed that many of the exposures of the verbal stimuli were at or above recognition threshold, some partial information was possibly available, and it could be argued that the subject elaborated this fragment of information into a general description of the face. The experimental effect was presumably slight because the stimuli were exposed only briefly (though repeatedly).

If a brief and impoverished stimulus produces only a primitive and allusive response, it is not surprising that further decreasing the amount of input will lead to a gross autonomic response. Such a finding was reported by Lazarus and McCleary (1951) in a study which showed that PGR measures were sensitive to the stimulus in the absence of verbal report. If the amount of information is increased, on the other hand, the amount of effect upon subsequent responses should increase, and this assumption is supported by Bressler (1931) who showed that the extent of the effect varied directly with the intensity of the stimulus. In short, one can postulate that increments in stimulus intensity lead to increments in adequacy of response and we will label this the intensity argument. It makes two general assumptions:

Assumption 1. A response based on partial cues is on the same continuum with a response to a fully developed stimulus, but the former is a paler and less precise copy of the latter. The notion of one continuum has been discussed recently by Weiner and Schiller (1960).

Assumption 2. Awareness of the stimulus can be used to *gauge* the amount of information being registered, so that when awareness is minimal (as in the subliminal studies) only partial cues are being received and only a primitive and inarticulate response can result. Increased awareness is a sign that more information is being presented which improves the "fit" between stimulus and response.

The intensity argument accounts for the main effects of the subliminal studies with-

[1] The study was carried out as part of a research project supported by grants from the United States Public Health Service.

We are indebted to Charles Cofer, Robert R. Holt, George S. Klein, Fred Pine, and I. H. Paul for their many helpful suggestions, and are especially grateful to Earl Reynolds, formerly associated with the School of Commerce, New York University, for making available his classes and classrooms.

out postulating a process below awareness, and in the minds of some (Eriksen, 1960; Weiner & Schiller, 1960) this is a victory for parsimony. But one finding, which has appeared repeatedly, does not fit this argument —the finding that the effect of an impoverished stimulus varies inversely with its intensity. This finding first appeared in the study by Smith et al. (1959) in which the verbal stimuli were exposed at progressively longer exposures, descriptions being obtained after each exposure. Effects of these words on the ambiguous face were much stronger and more systematic at low levels of exposure than at high levels. An analogous finding was reported by Paul and Fisher (1959): their stimulus had a greater effect on images when the level of exposure was considerably below the subject's threshold than when it was just below the subject's threshold. Eagle (1959) reports that subjects most influenced by subliminal stimuli had the highest thresholds. Lapkin and Lippmann (1959), report a correlation of .57 ($p < .05$, one-tailed test) between the degree to which the stimulus exposure was below threshold, and the magnitude of its effect on the response measure. The implication is that as information (exposure level) is decreased the effect increases, in contradiction to the first assumption of the intensity argument; and in contradiction to the second assumption, we find that awareness cannot be used to predict the amount of effect.

These findings seemed to embarrass the intensity argument, which uses awareness as an *indicator* of registration, and they provided the impetus for another approach. First advanced by Freud (1953) and later elaborated by Klein et al. (1958) and Fisher (1957), this approach makes four important assumptions: registration of the stimulus is independent of awareness, with awareness and registration conceptualized as separate aspects of cognitive activity; degree of awareness of a stimulus at the moment of input cannot be used to predict the extent of its effect on a response; the effects of a faint and weak stimulus are not necessarily weaker than those of a strong and clear stimulus; and reduction of awareness changes the patterning of cognitive activity.

Where the intensity argument implies that responses diminish to zero as the threshold is approached, the second argument states that a new class of responses is produced as we go below threshold. (We are not talking here of a physiological threshold. Obviously there is a lower limit below which no responses are possible.) The germ of this argument was set forth in the *Interpretation of Dreams* where Freud (1953) noted that thinking below awareness tends to fan out over a large network of traces.

> The energy attaching to the train of thought is diffused along all the associative paths that radiate from it; this energy sets the whole network of thoughts in a state of excitation which lasts for a certain time and then dies away . . ." (p. 594).

Freud also assumed that the associations in this network were logically organized and followed the rules of the secondary process. Putting these assumptions together we find that reduction of awareness should not necessarily lead to a more primitive response but to a greater spread of activity to other logically related responses. One idea can spread more freely throughout the system than is possible in conscious thinking where it is usually restricted to one train of thought; one percept can make contact with a wider range of traces. Support for this assumption is provided by experiments by Lacey and Smith (1954), Spence (1961), and Fox (1960), all of which show that, when not in awareness a verbal stimulus may generalize to a broad range of meanings. The generalization is often greater than is found with a conscious stimulus, and this finding suggests that when awareness is reduced, certain restrictions on thought are removed; more rather than fewer associative pathways become available and thought is freer to range over a greater number of associates.

Suppose a test word is exposed below threshold and followed by a list composed of words that are either related or unrelated to the test word. Subjects are then asked to recall the word list. Suppose another group of subjects undergoes the same procedure except that the test word is presented above awareness. The intensity hypothesis would predict that the subliminal group, receiving only a briefly exposed stimulus, could respond to it only on

the basis of partial cues—an isolated letter, for example—and would therefore tend to recall list words composed of the same, or similar, letters. These effects would vary with awareness; subjects who could barely discriminate the test word would get less information and would show a more random recall than subjects with a greater number of correct discriminations. If the latter group received enough information they might apprehend the meaning of the stimulus and recall related words. Thus the number of correct discriminations should be positively correlated with recall of related words. The supraliminal group, fully aware of the test word, also should respond in terms of meaning and, because it gets more information than the first group, should remember more words on the list which were meaningfully related to the stimulus.

But from our alternative hypothesis we would draw opposite predictions. We would assume that the subliminal stimulus would make contact with a wide range of traces and facilitate the recall of many associated words. Because thinking tends to fan out when it goes on outside of awareness (awareness, we propose, has a restricting effect), the spread of effect will be *inversely* correlated with amount of the stimulus in awareness. Thus subjects with few correct discriminations of the test word would be expected to recall more of the meaningfully related list words than would subjects with a high number of discriminations, and subjects in the Subliminal group would recall more such words than subjects in the Supraliminal group. The influence of partial cues would be relatively slight, subordinated to the influence of meaning, and not perfectly related to amount of awareness.

METHOD

If the meaning of a word registers outside of awareness and fans out along associative pathways, it would be likely to influence other associates which can be determined by word association norms. From the Minnesota norms recently collected by Russell and Jenkins (1954), we selected the word CHEESE as our stimulus and 10 associates to CHEESE which range from popular (MOUSE) to rare (BRICK). We chose words which appeared to be minimally interrelated to prevent the recall of one word from triggering off the others; then the number of words recalled could be used as a measure of size of effect. We also selected 10 control words which were not listed as associates to CHEESE, but which matched the CHEESE associates in Thorndike-Lorge (1944), frequency, number of letters, and part of speech. The two sets of words are presented in Table 1.[2]

TABLE 1

SETS OF STIMULUS AND OF CONTROL WORDS

Minnesota frequency[a]	CHEESE associates	T-L frequency[b]		Control words
94	MOUSE	34	32	hedge
82	BREAD	A	A	flag
32	COTTAGE	46	46	mirror
21	SMELL	A	A	gift
19	MOON	AA	AA	best
18	GREEN	AA	AA	chair
7	COW	A	A	sand
6	SOUR	15	16	frail
2	CAVE	33	34	arch
1	BRICK	49	48	trunk

[a] Responses per 1,008 subjects.

[b] Occurrences per million words. A = 50 per million; AA = 100 or more per million. Words were presented in the following order: street/bird/dime/flag/COW/arch/BREAD/CAVE/chair/MOON/best/COTTAGE/gift/frail/GREEN/sand/BRICK/hedge/SMELL/MOUSE/mirror/SOUR/trunk/chin/number/joke.

The first and last sets of three words are to control for primacy and recency effects and do not enter into the analysis of the findings.

Further support for our assumption that CHEESE associates and control words are equally balanced for recall can be noted in the finding that of 35 subjects so far tested in a blank condition (19 from this study plus 16 from another study), the mean recall of CHEESE associates (2.83) was not significantly different from the mean recall of control words (2.57).

If meaning is an important factor in recall, more CHEESE associates should be recalled than control words. If, on the other hand, fragments of a word register instead of its meaning, recall of structurally similar words should exceed recall of structurally different words. To test this hypothesis, the stimulus word CHEESE was divided into the two-letter combinations of ch, he, and es and the control words divided into a group which contained these letter-pairs (*ch*air, ar*ch*, *he*dge, and b*es*t) and those which did not (gift, flag, mirror, and trunk). If partial information influenced recall, the first group would be recalled better than the second group (none of which contained a single letter in common with CHEESE).

[2] Subsequent data have supported our assumption that the words are minimally interrelated. As part of another study, 70 subjects gave word associations to each of the 26 words on the recall list. One CHEESE associate produced another CHEESE associate as a response in only 5 cases out of 700; one control word produced another control word as a response in only 2 cases out of 700. The interitem associative strength, to use an expression coined by Deese (1959), is, therefore, essentially zero and there is little reason to assume that stimulus words cluster any more readily in recall than control words.

Procedure

Our design called for two groups equated for associative tendencies, and to obtain the relevant data we presented a large class with a list of 25 words taken from the Minnesota norms. Jenkins and Russell (1958) report that the number of popular associations to these 25 words is highly correlated with the number of populars given to the entire list of 100 words. Associative preference (commonality) was determined for 48 subjects and they were divided into two groups of equivalent mean commonality. One group was assigned to the Subliminal condition, and the other to a Blank condition.

Blank and Subliminal conditions. Two days after the selection procedure the experimental procedure was carried out. The Subliminal group was assigned to seats facing the front of the room and arranged in five equal rows of equal mean commonality. The front row was situated 21 feet from the screen; at that distance, the visual angle of the stimulus was 7°. The back row was 33 feet from the screen, with a visual angle of 4.3°. Subjects in the Blank group were seated around the edge of the room and faced away from the screen.

The experiment consisted of the exposure of the test word (CHEESE), a single reading of the word list, recall of the word list, and a measure of threshold. Subjects in the Blank group were told to close their eyes, while subjects in the Subliminal group were asked to look at a fixation point on the screen, and were told that something would be flashed in the neighborhood of that point. They were to write down what they thought it was, and how many times it flashed. The projected stimulus extended 1 foot 3 inches on either side of the fixation point. It was flashed five times at 1/150th of a second from a slide projector with an Ilex shutter. Room illumination was maximized, and the shutter diaphragm was set at its smallest opening. The combination of high room illumination, low stimulus intensity, and brief exposure worked effectively to prevent conscious recognition. No subject reported seeing anything, and the general reaction of the group was that nothing had been flashed. Estimates of the number of exposures varied from 4 to 5, but this seems to have been based on the number of shutter clicks heard by the subjects.

The Blank group was then asked to open their eyes, face front, and both groups were given the following instructions:

> We're going to read a list of 26 words. Later on we will ask you to reproduce them from memory. This is not a test, for it is quite impossible to remember all the words. Just relax and listen closely while we read the list. We will read it through once.

The list was then read by an instructor who had no knowledge of the purpose of the experiment. Three "padding" words were read first to absorb the effect of primacy; then the 10 CHEESE associates and 10 control words were presented in scrambled order; then three more "padding" words were added to control for recency. The complete list is given in Table 1.

Immediately after the list was read, subjects were told to write whatever words they could remember, in whatever order, and to continue writing for as long as they wished. They were encouraged to guess.

Threshold. To check on the subject's degree of awareness of the test word, we followed a suggestion advanced by Goldiamond (1958) and presented the stimulus and a blank slide in a series of 20 exposures in a fixed random order. Each slide was presented under the same condition of illumination and exposure as were used in the main part of the experiment. If any subjects had correctly identified CHEESE they would have been eliminated from the group. Subjects were given these instructions:

> Watch the screen once again, at the dot. Sometimes we will show you something very faintly and hard to see. Other times we will show you nothing. After each flash, write something or nothing. If you saw anything, please describe it at the bottom of the page.

The number of correct discriminations provided a measure of the relative exposure level of the stimulus.

Supraliminal condition. Associative commonality was determined for a separate class using the procedure described above; 2 days later subjects were arranged in front of the screen in such a way that each row had the same average commonality, and the distance of each row from the screen matched the distance for the Subliminal group.

Subjects were told to expect something to be flashed on the screen. They were told to watch the fixation point and write down what they saw, and how many times it flashed. CHEESE was exposed 5 times in exposures of 2 seconds each. Its projected size and placement on the screen were the same as in the previous Subliminal session. The group was then read the same preparatory instructions used in the first experiment; the standard list was read by the same instructor, and subjects wrote down their recall. All subjects perceived the stimulus CHEESE, and all but one counted five flashes (he counted four).

Subjects

The members of all three groups (Subliminal, Supraliminal, and Blank) were taken from the undergraduate classes of the School of Commerce, New York University. Of the 65 subjects, 59 were male. Twenty-nine subjects composed the Subliminal group, 19 the Blank group, and 17 the Supraliminal group. The Subliminal group was made purposely larger than the rest in order to better investigate the phenomenon. All three groups were taught by the same instructor and the experiments took place during regular class sessions. All subjects and the instructor were naive as to the purpose of the experiment; no subjects were majoring in psychology.

RESULTS

Effect of Subliminal Stimulus on Recall

The stimulus CHEESE was not consciously perceived by any subjects in the Subliminal group. Six of the 29 subjects did detect the presence of "something" in the discrimination series significantly better than chance ($p < .05$, one-tailed test), and they were eliminated from the main analysis in the interest of caution. The recall of the remaining 23 subjects was analyzed for effects of structure and meaning of the subliminal stimulus.

Structure. If the structure of the stimulus CHEESE influenced recall, more structurally related control words (*ch*air, ar*ch*, *h*edge, and be*s*t) should be recalled than dissimilar control words (gift, flag, mirror, and trunk). There is a trend in this direction but it is not significant ($.10 > p > .05$, two-tailed, Wilcoxon test). However, if we add the six subjects who were able to discriminate the stimulus better than chance, the difference reaches significance ($p < .01$, two-tailed, Wilcoxon test) because these six subjects were responding primarily on the basis of structural cues. Their recall of the structurally similar words ranged from 1 to 3, with a mean of 2.2; their recall of the dissimilar words ranged from 0 to 2, with a mean of .33. When partial cues are available to the Subliminal group they facilitated recall of structurally similar words, as the intensity hypothesis would predict; *but this is only true at the level of partial awareness.* Below this level these cues have no significant effect, and their effect does not vary with intensity or awareness of the stimulus. Using only the 23 "chance discriminating" subjects, we computed each subject's recall of structurally similar minus dissimilar words; this net recall is correlated −.13 (*ns*) with distance from the screen—a measure of size and intensity of the stimulus—and −.28 (*ns*) with number of discriminations—a measure of awareness. Thus the intensity argument is not supported at subliminal levels of exposure.

For the Blank group, the difference in recall of the two groups of structurally related words was not significant.

Meaning. If the meaning of the CHEESE stimulus influenced recall, more CHEESE associates should be recalled than nonrelated control words. Nineteen of the 23 subjects in the Subliminal group recalled more CHEESE associates than control words ($p < .001$, two-tailed, Wilcoxon test), a strong indication that meaning had an effect.[3] Only 9 of the 19 subjects in the Blank group recalled more CHEESE associates (*ns*). These findings strongly suggest that the meaning of a stimulus that lies outside the range of correct discrimination can influence the recall of related words. The strong subliminal effect is presented in Table 2.

TABLE 2

MEAN RECALL OF CHEESE ASSOCIATES AND CONTROL WORDS UNDER THREE CONDITIONS

Condition	*N*	CHEESE associates	Control words	Difference[a]
Subliminal	23	4.09	1.91	2.18[b]
Blank	19	2.89	2.26	.63
Supraliminal	17	3.00	2.41	.59

[a] Differential recall is significantly higher in Subliminal than in Supraliminal or Blank conditions ($.01 > p > .001$, Kruskal-Wallis test). Differential recall is significantly higher in Subliminal than in Blank ($p < .02$, two-tailed, Mann-Whitney *U* test); significantly higher in Subliminal than in Supraliminal ($p < .02$, two-tailed, Mann-Whitney *U* test); and not significantly different between Blank and Supraliminal.

[b] Significant at .001 level. Significantly more CHEESE associates than control words are recalled in the Subliminal condition (two-tailed, Wilcoxon test). Recall between CHEESE and control words is not significantly different in the other two conditions.

The associates recalled by the 23 subject Subliminal group cover the full range of response frequencies. We computed the percentage of subjects in the Subliminal group who recalled each CHEESE associate and the rank-order correlation between these frequencies and Minnesota frequencies is .42 (*ns*). Since some of the recall probably depended on uneven stresses in pronunciation, uneven pausing, and other characteristics of the manner in which the list was read, we computed comparable percentages for subjects in the Blank group, and subtracted one set of frequencies from the other. Since both groups listened to the same list, any difference in re-

[3] A replication has recently been completed by the senior author. Nineteen subjects received the subliminal stimulus and 16 subjects served as controls, using the same procedures described above. Recall of CHEESE associates minus control words was significantly better in the Subliminal than in the Blank group ($p < .06$, one-tailed test).

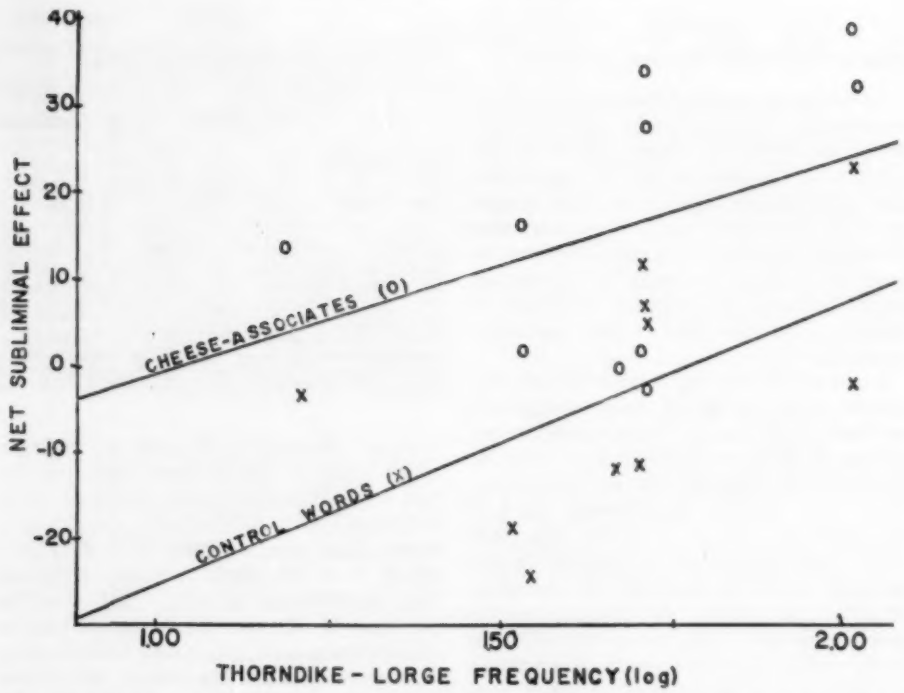

FIG. 1. Net subliminal effect (percentage of recall in the Subliminal group minus percentage of recall in the Blank group) as a function of Thorndike-Lorge frequency. (Regression lines as indicated.)

call was presumably due to the subliminal stimulus and was free of "list effect." Although these differences are likewise not correlated with Minnesota frequency (rho = − .10), they are strongly related to Thorndike-Lorge word frequency (rho = .63, $p <$ .05, two-tailed test).[4] Within the group of CHEESE associates, the extent of recall is determined more by the general popularity of each word than by its popularity as an associate to CHEESE.

A correlation with word frequency is also reflected in the net recall of control words (rho = .78, $p < .02$). Apparently the frequency of occurrence determines the readiness of recall; words that are infrequently used in written English are unfamiliar and less likely to be recalled. If a word was meaningfully related to CHEESE, it received an additional increment which seems to be more or less independent of its association strength (see Figure 1 in which the gradient of recall for the CHEESE associates is roughly parallel, but shifted above, the gradient for the control words). The likelihood of recall for a particular word appears to have been the sum of its initial habit strength plus a more-or-less uniform subliminal reinforcement.

However, when word frequency is partialed out, we do find an interesting relation to associative strength. Amount of net recall of the five most popular CHEESE associates (A or AA words on Table 1) is correlated − .90 ($p < .10$, two-tailed test, rank-order correlation) with Minnesota frequency which shows that the subliminal stimulus has a stronger effect on its weaker associates, other things being equal.

Since recall was influenced by meaning, we next asked whether it was influenced by the subject's associative preference (commonality); it was not. Although subjects with high

[4] We are indebted to George S. Klein for suggesting the existence of this relationship.

commonality tend, by definition, to give popular responses on an association test, they did not tend to recall popular CHEESE associates. The average association strength of the associates recalled by subjects is not correlated with commonality (rho = − .06). These results show that the subliminal stimulus reinforces a broad network of associates which are not concentrated around the most preferred conscious response and (to judge by the preceding paragraph) tend to be distantly related to the stimulus. Apparently the subliminal stimulus follows laws that are independent of those which determine conscious associations.

Influence of Awareness on Recall

Recall of CHEESE associates in the Supraliminal condition was not significantly greater than control words (see Table 2) which indicates that exposure of an organizing word does not necessarily lead to an increase in its associates. This finding parallels that of Deese (1959) who found that recall of associates was not increased when the list was preceded by the stimulus for those associates. Moreover, we find that differential recall of CHEESE associates is higher in the Subliminal than in either the Blank and Supraliminal conditions, and differences between conditions are highly significant ($.01 > p > .001$, Kruskal-Wallis one-way analysis of variance). Table 2 clearly shows that the significance stems from the high difference in the Subliminal condition, and subsequent Mann-Whitney tests reveal that differential recall is significantly greater in Subliminal than in Supraliminal ($p < .02$, two-tailed test) and greater in Subliminal than in Blank ($p < .02$, two-tailed test). The difference between Blank and Supraliminal is clearly not significant. Apparently the greatest effect on recall of CHEESE associates is produced by the subliminal stimulus; comparing subliminal and supraliminal effects, we find that awareness of the stimulus apparently restricts its effect on recall. If this is generally true, those subjects in the Subliminal group who were maximally aware of the stimulus should recall fewer CHEESE associates than those subjects who had minimal awareness of the stimulus. To test this assumption, we replaced the six subjects who had been dropped earlier from the Subliminal group and computed the correlation between number of discriminations, which ranged from 8 to 19 out of 20, and relative recall (CHEESE associates minus control words). The rank-order correlation is − .54 ($p < .02$, two-tailed test); poorer discrimination is accompanied by greater net recall of words which are meaningfully related to CHEESE (see Figure 2). As discrimination improves the effect falls off.

In view of the greater effect of partial cues among high discriminating subjects noted earlier, and the greater effect of meaning among the randomly discriminating subjects, we made a direct comparison of the two factors of structure and meaning. We compared recall of arch, chair, best, and hedge, the four structurally similar control words, with four CHEESE associates of equal word frequency but a lower degree of structural similarity—CAVE, GREEN, MOON, and MOUSE. Two of these words contain pairs of letters which are identical to a pair of letters in the test stimulus; one has similar letters not contained in pairs; and the fourth has no similar letters. We found that subjects at the low end of the discrimination continuum recalled more CHEESE associates than control words; that subjects at the high end recalled more control words than CHEESE associates; and that, overall, there is a correlation of .57 ($p < .02$,

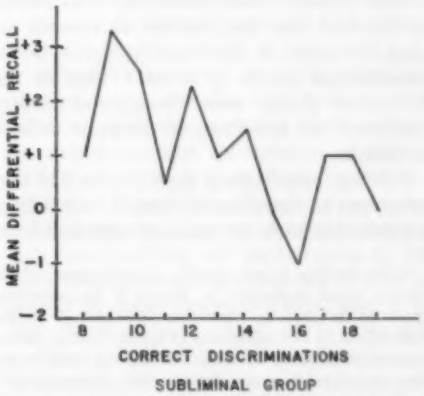

FIG. 2. Mean differential recall (CHEESE associates minus control words) as a function of correct discrimination in the subliminal condition.

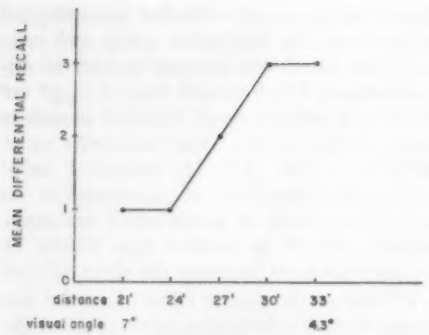

FIG. 3. Mean differential recall (CHEESE associates minus control words) as a function of distance and visual angle in the subliminal condition.

two-tailed test) between number of discriminations and the preference for control words. Evidently structural cues gave way to meaning at low levels of awareness, but began to organize recall at a high level of awareness. These findings suggest that partial cues, as they became more available, tended to organize the recall in terms of structural elements and thus reduced the recall of meaning associates (which have less structural correspondence to the test stimulus).[5]

Another kind of limiting effect can be seen when we study the intrusions—words which were "remembered" but were not on the original list. The supraliminal stimulus tended to direct attention toward food associates to CHEESE (sweet, taste, meat, etc.) as shown by the fact that the number of CHEESE related intrusions is significantly higher in the Supraliminal group ($p < .001$) than in the Subliminal group, even though the average number of all intrusions is the same in both conditions.

Putting these findings together, we find that awareness of the stimulus tended to reinforce a particular area of meaning and restricted sensitivity to other connotations. Overemphasis on one category of meaning markedly reduced the range of words reinforced by the perceived word CHEESE and led to the paradoxical finding that a supraliminal stimulus had a more restricting effect on recall than a subliminal stimulus.

Influence of Intensity and Size on Recall

Our subjects in the Subliminal condition were arranged in five rows of varying degrees of distance from the screen, and by taking advantage of the fact that intensity and size vary with distance, we had a chance to measure how these variables influence effect. Recall of CHEESE associates minus recall of control words was positively related to distance from the screen (rho = .63, $p < .02$, two-tailed test; see Figure 3). At the maximum distance, the visual angle (4.3°) is only slightly greater than the foveal angle and thus the most sensitive part of the retina was being stimulated by the test word.[6] At shorter distances less of the word was impinging on the fovea, in which case less of its meaning would be communicated and we would expect that only single letters of the test word would have an effect on recall. Using the same two groups of words mentioned earlier, we found that subjects in the front rows recalled more structurally similar control words (ar*ch*, *ch*air, b*e*st, and *h*edge) than the more dissimilar CHEESE associates, (CAVE, GREEN, MOON, and MOUSE); that subjects in the back rows recalled more dissimilar CHEESE associates than similar control words; and that, overall, preference for meaning vs. structure as an organizing principle was correlated .64 with distance ($p < .02$, two-tailed test, rank-order correlation). Apparently there was a tendency for subjects in the front rows to react more in terms of single letters than in terms of the entire word, and this is consistent with the assumption that only parts of the test word were effectively registering.[7]

[5] This finding is not simply a restatement of the general trend expressed in Figure 2 for meaning-effect to decrease as awareness increases. A similar comparison of two other sets of words (BREAD, SMELL, COTTAGE, and BRICK vs. flag, gift, mirror, and trunk) does not show any correlation with awareness (rho = .25); i.e., the high discriminating subjects are not less sensitive to meaning when the equivalent control words (flag et al.) have no structural similarity to the test stimulus.

[6] We are indebted to Leo Hurvich for pointing out this possibility.

[7] Closeness to the screen also seems to restrict the recall of meaning in the absence of partial cues. Recall of BREAD, SMELL, COTTAGE, and BRICK minus flag, gift, mirror, and trunk is correlated .45 with row ($p < .05$, two-tailed test). Flag et al. have no struc-

Distance from the screen is, surprisingly, correlated only .15 (rho) with amount of discrimination. Thus we are justified in saying that a decrease in distance did not automatically lead to awareness of the stimulus, and that the factors which contribute to broad effects in the back rows are probably independent of the factors which contribute to good discrimination. The "row effect" probably depends on such general factors as size, intensity and visual angle; the "discrimination effect" on individual factors such as visual acuity, readiness to guess, distribution of attention, etc. Obviously these two sets of factors can interact in a wide variety of ways.

DISCUSSION

The results show that the meaning of a stimulus can be registered without awareness and can significantly influence subsequent recall. The pattern of recall shown by the Subliminal subjects makes it more parsimonious to conclude that the word CHEESE was registered intact and undistorted, and does not support the assumption that some similar structure was registered which bore only an approximate resemblance to CHEESE. A similar, but not identical, structure would not explain why meaningfully related associates were recalled.

The second conclusion to be drawn from our findings is that reported awareness is not a reliable indicator of the amount of information being registered. Subjects who were unable to discriminate the presence or absence of the test word nonetheless received a substantial amount of meaning information, and seemed relatively uninfluenced by partial cues. Those subjects who could discriminate the stimulus, however, were sensitive to partial cues, and were not sensitive to meaning. If anything, reported awareness is an indicator that more structurally determined effects are to be expected and the absence of awareness, an indicator that more meaning determined effects are to be expected, as shown by the correlation between degree of awareness and preference for structure over meaning in the recall.

The results also fail to support the assumption that the response to reduced stimulus input is a crude and pale copy of the response to more intense input. The responses of subjects in the back row of the Subliminal group, receiving the stimulus at substantially reduced intensity, were no less dependent on meaning than were the responses of subjects in the front row. And when information was substantially increased by lengthening exposure speed, the accompanying responses did not include a greater number of CHEESE associates. Subjects in the Supraliminal group received the stimulus about 300 times longer than subjects in the Subliminal group, but the response of the former group was no more complete, as would be reflected in a greater recall of CHEESE associates.

These facts lead us to conclude that the intensity argument is insufficient to account for the results, and take us back to our alternative set of four assumptions which more nearly meet the case. Our findings support Assumption 1 that registration takes place without awareness; Assumption 3 that a subliminal stimulus does not have a weaker effect than a supraliminal stimulus; and Assumption 4 that reduction of awareness changes the patterning of cognitive activity, with different associative processes operating in the two conditions. Assumption 2 no longer holds, and should be rephrased to say that awareness is inversely related to amount of effect.

Before we consider the apparent paradox that an unseen stimulus may have a stronger effect on recall than a clearly visible stimulus, we might consider the pattern of the subliminal effect. It is clear that a reduction of stimulus information does not necessarily break down the structure of what is registered and leave nothing but partial cues to influence the response. Meaning can still be preserved at faint exposure levels, and it fans out along a wide range of predetermined pathways. The pattern of nonconscious generalization does not seem to be influenced by the subject's conscious associations (his commonality) and—rather surprisingly—seems to be inversely related to the Minnesota norms.

tural similarity to the test word. Apparently, the closer the subject is to the screen, the fewer letters of the test word impinge on his fovea and the less meaning is conveyed, thus restricting recall.

When Thorndike-Lorge frequency is held constant, the activation of rare associates seems stronger than of the more popular associates. For some reason the subliminal effect on the popular associates is inhibited and, perhaps as a consequence, it fans out to influence other, more distant associates.

It is also worth mentioning that the subliminal effect, although relatively indirect, is not chaotic. Clearly the test word did not activate a host of unrelated meanings, for if it had there would have been no measurable effect. Activation is directed primarily to words related to the stimulus, and conscious attention is not necessary to keep simple thought processes under control. Preconscious thinking is not only more than a theoretical possibility, but seems to follow the laws of the secondary process, albeit in an inefficient way, for the fanning-out effect, by its very nature, would certainly interfere with the sustained, secondary process thought necessary for solving a particular problem.

This fanning-out effect seems to be greatest in the last two rows of subjects (see Figure 3) where the stimulus falls almost entirely within the acutely sensitive part of the retina. Since all parts of the word must be registered to convey its meaning, the transmission of meaning is more likely when there is complete foveal impingement. As the distance decreases, the number of letters registered in the fovea also decreases and meaning is correspondingly impoverished, probably because the letters which are registered foveally play a disproportionate part in influencing the response, and may bend it away from the category of CHEESE and CHEESE associates and toward words which have only a structural similarity to the test word. In short, the "row effect" may be due to the fact that the fovea is stimulated by the entire word—which conveys meaning—in the back rows, and by single letters—which produce other response alternatives—in the front rows. The correlation between row and tendency to recall meaning associates seems to support this assumption.

So much for distance; what about awareness? Subjects who discriminated better than chance were more sensitive to partial cues than to meaning, a fact which supports the assumption that single letters somehow inhibit the activation of CHEESE associates. Partial information, since it can be elaborated in a variety of ways, may lead to hypotheses which have little to do with CHEESE and its associates and hence will not improve recall. Fragments of a word seem to force recall of structurally similar words over CHEESE associates with fewer structural similarities. As awareness increased, partial information also increased and tended to restrict the effect on recall.

Thus it would seem that both awareness and wide visual angle may favor registration of single letters over the complete word, and thus lead to a variety of individual hypotheses based on these letters. One might posit three different stages of registration depending on intensity and size of the stimulus. In Stage 1, the stimulus is well below threshold, excites a wide band of association pathways, and produces a broad activation of many conceptual associates. In Stage 2, bits of partial information play a disproportionately greater role and generate hypotheses based on single letters which organize recall. Since these hypotheses can assume a variety of forms, many of which are quite unrelated to the stimulus, they may easily interfere with the expected effect. In Stage 3, the stimulus is supraliminal, clearly within awareness and organizes the response in terms of a particular area of meaning, as shown by the fact that subjects in the Supraliminal condition reported more food associates than subjects in the Subliminal condition. Note that the effect is restricted in Stage 3 for quite different reasons than in Stage 2.

The part played by partial information in Stage 2 may account for the paradox noted earlier in the paper—the fact that the effect of a verbal stimulus appears to diminish as it reaches threshold. In the study by Smith et al. (1959), a subject might elaborate the stimulus ANGRY into angel as it approached threshold, and describe the face in positive terms. Any increase in idiosyncratic responses such as this would tend to dilute the expected effect, and it is important to note that individual parameters are most likely to organize the response as soon as bits of information begin to emerge into awareness. Partial information, for this reason, may be more mislead-

ing than no information at all, and reports based on partial information may be erroneous not simply because less information is presented, but because this information is organized around an idiosyncratic premise that may be only distantly related to the stimulus.

This, then, is the restricting effect of awareness and this is why a subliminal stimulus, well below the threshold of partial recognition, has a more widespread and uniform effect. Randomly discriminating subjects in the Subliminal group remembered more CHEESE associates; Spence and Bressler (in press) found that the effects produced by a stimulus far below threshold conformed more exactly to the Minnesota norms than the effects produced by a stimulus closer to threshold. In both cases, as the stimulus moved closer to threshold it gave rise to individual response tendencies (probably mediated by partial cues) and the accompanying response was stamped by idiosyncratic earmarks.

In view of these findings, we must take issue with Eriksen's (1960) recent statement that "there is no convincing evidence that the human organism can discriminate or differentially respond to external stimuli that are at an intensity level too low to elicit a discriminated verbal report" (p. 298) and say that a stimulus which is not verbally discriminated is nevertheless capable of remarkably accurate and far-reaching effects which involve meaning and stand in contrast to the structural effects mediated by partial cues.

SUMMARY

The purpose of this experiment was to study how level of awareness influenced the effect of a verbal stimulus on recall. Sixty-five undergraduate students were presented a list of words to recall, half of which were associates to the test stimulus CHEESE, and half were not. Before the list was read, the stimulus was presented subliminally to one group, supraliminally to a second, and never exposed to the third, Blank group. Whereas both Blank and Supraliminal groups showed no preferential recall of CHEESE associated words, the Subliminal group showed a significantly greater recall. Degree of effect was inversely related to awareness of the stimulus, and to closeness of the stimulus. Further analysis of the role of level of awareness led to the postulation of three stages of registration: Subliminal with no partial cues; Near-liminal with partial cues available; and Supraliminal. More meaning associates were recalled in Stage 1; more structurally related words in Stage 2. The features characterizing each stage were discussed in detail and led to the conclusion that an explanation based on partial cues may explain Stage 2, but not Stage 1 for which additional assumptions are necessary.

REFERENCES

BRESSLER, J. Illusion in the case of subliminal visual stimulation. *J. gen. Psychol.*, 1931, **5**, 244–250.

DEESE, J. Influence of inter-item associative strength upon immediate free recall. *Psychol. Rep.*, 1959, **5**, 305–312.

EAGLE, M. The effects of subliminal stimuli of aggressive content upon conscious cognition. *J. Pers.*, 1959, **27**, 578–600.

ERIKSEN, C. W. Discrimination and learning without awareness: A methodological survey and evaluation. *Psychol. Rev.*, 1960, **67**, 279–300.

FISHER, C. A study of the preliminary stages of the construction of dreams and images. *J. Amer. Psychoanal. Ass.*, 1957, **5**, 5–60.

FOX, MURIEL. Differential effects of subliminal and supraliminal stimulation. Unpublished doctoral dissertation, New York University, 1960.

FREUD, S. *The interpretation of dreams.* (Originally published 1900) Vols. 4 & 5. (Trans. by J. Strachey) London: Hogarth, 1953.

GOLDIAMOND, I. Indicators of perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychol. Bull.*, 1958, **55**, 373–411.

JENKINS, J. J., & RUSSELL, W. A. Miscellaneous studies in word association. Technical Report No. 24, 1958, University of Minnesota, Contract N8 onr 66216.

KLEIN, G. S., SPENCE, D. P., HOLT, R. R., & GOUREVITCH, SUSANNAH. Cognition without awareness: Subliminal influences upon conscious thought. *J. abnorm. soc. Psychol.*, 1958, **57**, 255–266.

LACEY, J. I., & SMITH, R. L. Conditioning and generalization of unconscious anxiety. *Science*, 1954, **120**, 1045–1052.

LAPKIN, B., & LIPPMANN, P. Is there a need for two concepts: Perception and registration? Unpublished manuscript, New York University, Research Center for Mental Health, 1959.

LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: A study of subception. *Psychol. Rev.*, 1951, **58**, 113–122.

PAUL, I. H., & FISHER, C. Subliminal visual stimulation: A study of its influence on subsequent im-

ages and dreams. *J. nerv. ment. Dis.*, 1959, **129**, 315–340.

PINE, F. Incidental versus focal presentation of drive related stimuli. *J. abnorm. soc. Psychol.*, 1961, **62**, 482–490.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract N8 onr 66216.

SMITH, G. J. W., SPENCE, D. P., & KLEIN, G. S. Subliminal effects of verbal stimuli. *J. abnorm. soc. Psychol.*, 1959, **59**, 167–176.

SPENCE, D. P. The multiple effects of subliminal stimuli. *J. Pers.*, 1961, **29**, 40–53.

SPENCE, D. P., & BRESSLER, JOANNA. Subliminal activation of conceptual associates: A study of rational preconscious thinking. *J. Pers.*, in press.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Bureau of Publication, 1944.

WEINER, M., & SCHILLER, P. H. Subliminal perception or perception of partial cues. *J. abnorm. soc. Psychol.*, 1960, **61**, 124–137.

(Received March 1, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 64, No. 3, 175–182

SOME EFFECTS OF REJECTION UPON ATTRACTION TO A GROUP [1]

J. DIEDRICK SNOEK

Research Center for Group Dynamics, University of Michigan

The process by which individuals move in and out of groups can be thought of as a mutual selection during which the individual evaluates the attractiveness of membership in a particular group and the group judges his acceptability as a member. Among the factors of interest to social psychologists in this process are the motivational basis for the individual's initial interest in becoming a member, the criteria by which the groups accept or reject him, and finally the basis for his continued participation in the group. While previous studies have paid considerable attention to the bases for an individual's attraction to membership, given his acceptance into a group, few studies have concerned themselves with changes in the nature of an individual's desire for membership in a group as a result of his acceptance or rejection. The study to be reported below concerns some conditions under which an individual will continue to be attracted to membership in a group which has rejected him.

The experience of rejection has two important psychological aspects. First, rejection deprives the individual of the opportunity to gratify whatever needs led him to desire membership in the group. Secondly, rejection represents an evaluation by the group of the rejected person. To the extent that the criteria used by the group in making its evaluation are revelant to the individual's self-evaluation, rejection may be more or less damaging to self-esteem. In this paper we shall speak of *invidious rejection* whenever exclusion from a group can be taken to mean that the individual is not worthy of membership in it. We shall call the rejection *non-invidious* when it is based on other reasons that do not reflect on the individual's self-esteem; when an individual is excluded because all membership positions in a group are filled we have an example of noninvidious rejection.

Just as it is possible to distinguish various degrees of acceptance in a group, various degrees of rejection are possible. Several experiments have tried to manipulate this variable, although conceptually it has not always been clear whether "degree of rejection" referred to the amount of frustration experienced by the rejected person or to the negativity of the evaluation he received. In this paper the concept of degree of rejection will refer to the strength of the group members' determination to resist the individual's attempts to participate as a member in the group, after Jackson's (1959) usage.

The concept of attraction to a group often is taken to refer to the strength of an individual's desire to participate in a group without regard to the particular needs that may motivate him. In this paper attraction will be considered a joint function of the valence of membership in the group and the expectation that membership can be attained, which can be expressed as the individual's subjective probability of acceptance. That is to say that in a choice situation the individual's tendency to choose membership in a group will be a function of both the apparent attractiveness of the group and his expectation that he can join it (cf. Feather, 1959). For convenience it is assumed that the function relating valence and subjective probability of acceptance is a simple multiplicative one.

Given this formulation of attraction, two hypotheses can be formulated regarding the expected effects of rejection upon attraction to a group. First, whatever the reasons for rejection, the individual will revise his estimate of the probability of his acceptance

[1] This paper is based upon a doctoral dissertation submitted to the Graduate School of the University of Michigan in 1959. The author wishes to acknowledge his gratitude to the Social Science Research Council for the fellowship which permitted him to conduct this experiment.

downwards to an extent that is proportional to the degree of rejection. The result of such a change would be a decrease in his overall attraction to the group. The second effect to be expected is a change in the valence of membership depending on whether or not the rejection was invidious. Jackson and Snoek (1959) found that the reactions to rejection of very strongly rejected subjects, as contrasted with those who only experienced mild rejection, appeared to be motivated by a need to reassure themselves about their acceptability to others (need for social reassurance). The strongly rejected subjects maintained their attraction to the group, while subjects in the mild rejection conditions appeared to have decreased theirs. In the present study it is expected that the need for social reassurance will be strongly aroused in invidiously rejected subjects who in turn will value reacceptance into the group as a means of reassuring themselves. Accordingly, the valence of membership is expected to increase when the reasons for rejection are invidious; this change in valence is assumed to be proportional to the degree of rejection.

In summary, invidiously rejected subjects should experience two opposing tendencies to change their attraction to the group, one to lower their attraction due to a decrease in their estimate of the probability of acceptance and a second tendency to increase their attraction due to their need for social reassurance. By contrast noninvidiously rejected subjects will be subject only to the first tendency to lower their attraction to the group. From these assumptions it is possible to predict the following differences in relative changes in attraction to a group after the two kinds of rejection:

Hypothesis 1. For any given degree of rejection, noninvidious rejection will be followed by a greater reduction in attraction to the group than invidious rejection.

Hypothesis 2. Differences between invidious and noninvidious rejection conditions in loss of attraction to the group will be greater after strong than after mild rejection.

METHOD

Subjects. Experimental groups were organized by asking male undergraduates at the University of Michigan to join a series of bridge clubs, organized for purposes of experimental observation as well as for their enjoyment. Fifty-seven subjects participated in the experiment. Subjects were both recruited and scheduled for meetings by the author, identified as a graduate assistant at the Research Center for Group Dynamics.

Procedure. The usual group consisted of two subjects and two hired experimental assistants, one male and one female, who pretended to be subjects. At the time of scheduling the experimenter told each subject he would be meeting with an already established group which lacked a fourth member; in this way subjects were put in the position of being candidates for the position of "fourth member" in an already existing group. Since subjects generally arrived at the meeting at slightly different times, it was possible to introduce each of them separately to the two assistants and then to each other, so that each subject remained unaware that the other was also a newcomer to the group.

After the introductions the members of each group were told to proceed as they wished, while the experimenter took the passive role of observer. Most groups immediately started to play bridge, using a set of predealt hands arranged so as to keep each subject's and group's bridge experiences comparable; groups which spent more time getting acquainted were prodded into playing bridge after a short time by one of the experimental assistants. After approximately 1 hour, the experimenter requested an interruption in the meeting in order to administer a questionnaire. For this purpose subjects were separated into cubicles, a procedure which allowed the experimenter to administer the subsequent rejection induction to each of the two subjects in private. After the rejection manipulation subjects rejoined the group for another 45 minutes of bridge playing, at the end of which they were again asked to fill out a questionnaire. After the second questionnaire, subjects were reassembled, told of the experimental manipulations and given a chance to share their reactions with the other participants.

Four experimental conditions were created: Strong Invidious, Mild Invidious, Strong Noninvidious, and Mild Noninvidious rejection. In all four conditions rejection was induced by the experimenter under the pretense of feedback of spurious questionnaire ratings subjects had made of each other. The following instructions were used:

> Because you are in the unusual position of being a newcomer to this group and because the evaluations you received look somewhat different from those of other members in this group, I thought I would discuss these ratings with you now rather than wait till the end of the session, as I usually do. Do you remember that all of you were asked to write on the first questionnaire whether you preferred to continue playing with each other in future meetings? As you can see [showing spurious evaluations] most of the people

in this group agree that they want to continue playing with each other.

[*Strong rejection treatment*] But as you can see, their rating of you is quite a bit lower than that of the others. In other words, most of the members of this group seem to prefer that you do not try to join this group.

[*Mild rejection treatment*] But as you can see, their rating of you is not quite as high as that of the others. In other words, the other members of this group seem to be neither in favor of nor opposed to your joining the group.

[*Invidious reason for rejection*] As the other ratings show, you are apparently considered not quite as good a bridge player as they would have liked; you are also not as well-liked as some others in this group. Apparently, they seem to think that you would fit rather badly into this group.

[*Noninvidious reason for rejection*] As the other ratings show, you are considered just as good a bridge player and just as likeable a person as anyone else in the group. What seems to be the case, however, is that they would prefer to have a second girl playing with them in the future, rather than having three men and only one girl as would be the case if you became the fourth member.

After these instructions subjects were enjoined to finish the questionnaire on which they were working and then to rejoin the group. Meanwhile the experimenter would repeat the rejection manipulation with the second subject in another cubicle. On a few occasions the second subject would already be out of his cubicle and no opportunity remained to give the rejection instructions without arousing suspicion; five subjects therefore received no experimental inductions. Four subjects questioned the rejection induction, stating that they thought it was an experimental manipulation; these subjects were eliminated from the analysis.

Questionnaire measures. Attraction to the group was measured with a set of three questions, taken from Libo (1953). Attraction scores based on summing over the three items showed a correlation of .92 (biserial *r*) with Libo's "locomotion criterion" of choosing to stay in or to leave the group (computed by the author on Libo's data). The three questions were: "Do you want to remain a member of this group?" (Desire to Stay), "How often would you like to come to meetings of this group?" (Frequency of Meeting), "If this group broke up for a considerable length of time and some people were trying to get it started again, would you want to rejoin?" (Desire to Rejoin).

A measure of Subjective Probability of Acceptance was obtained from answers to the question: "How likely do you think it is that the other members of this group want you to remain a member?" The question had five alternative answers, ranging from "very likely" to "very unlikely." Like the attraction questions, this question was repeated after the rejection manipulation and the second period of interaction with the group.

In addition, the final questionnaire included an item on Attraction to Membership in Another Group ("Instead of continuing to meet with this group, would you like to have the opportunity to meet with another group like this?") and one on Desire for Contact with individual members of the group outside of the experimental sessions, based upon a seven-item scale constructed by Back (1951).

Observation measures. Verbal interaction was recorded throughout the two periods of interaction, before and after the rejection manipulation. A record was kept of (*a*) the frequency with which each subject spoke, (*b*) the content of his remarks in terms of a very simple classification schema, and (*c*) the passage of time. The unit of speech coded was a *remark,* defined as a series of phrases or sentences, uttered without interruption, about a single topic. If the topic was changed once or more often during a single utterance, the subject was recorded as having initiated two remarks. Content was coded as either (*a*) *task demanded,* defined as remarks that were unavoidable in the context of the bridge game, such as bids or responses; (*b*) *task related,* defined as spontaneous remarks related to but not required by the bridge game; (*c*) *personal,* defined as remarks about the player himself or about one of the other players' background or attributes; or (*d*) *other,* a residual category. Pretesting showed that the great majority of remarks during the bridge game could be coded reliably into the first three categories. The reliability of the coding was estimated by correlating the rates of participation for each content category and participant obtained by two observers over nine observation periods of approximately 10-minutes duration. The obtained correlation coefficient was .86.

Role of experimental assistants. The two assistants played no role in the rejection manipulation, but were used only because it was impossible for the experimenter to deal with more than two subjects at a time. Their familiarity with the room and procedures reinforced the subjects' impression of joining an already "established" group. Because subjects for each group were selected on the basis of self-ratings of their bridge competence, the assistants were instructed to simulate a level of competence comparable to that of the subjects. The competence ratings received by the assistants showed they were indeed always considered to be roughly equal in competence to the subjects. Similarly, the assistants were instructed to initiate as little interaction as possible on their own, but instead to imitate as closely as possible both in content and volume the patterns of interaction their partners (the subjects) seemed to prefer, so that the interaction rates might reflect as closely as possible the subjects' preferences. Analysis showed that the assistants were reasonably successful in this attempt; their participation rates over all experimental sessions correlated .40 and .66 with their respective partners' scores.

TABLE 1

NUMBER OF SUBJECTS CHANGING PROBABILITY OF ACCEPTANCE ESTIMATES AFTER REJECTION

Probability of acceptance	Rejection conditions			
	Mild invidious	Mild noninvidious	Strong invidious	Strong noninvidious
No change	8	7	4	4
Decreased	4	5	8	8

Note.—Mild vs. Strong: $\chi^2 = 4.16$, $p < .05$; Invidious vs. Noninvidious: *ns*.

RESULTS

Due to the loss of 4 subjects who discovered the deception and 5 subjects who never received the rejection treatment, the analysis concerns 48 subjects who were randomly assigned to four experimental conditions, 12 to each condition.

The effectiveness of the strength of rejection instructions could be checked by comparing the relative decrease in subjective probability of acceptance for the mild and strong rejection conditions. Table 1 shows that the strong rejection instructions led to a reliable increase in the number of subjects who decreased their estimate of the likelihood of being accepted by the group. Reason for rejection, as expected, had no discernible effect upon changes in subjective probability of acceptance.

Prior to the experiment subjects returned a mailed questionnaire, consisting of 20 seven-point scales, on which they had described the kind of group they would prefer to join. After the first interaction period subjects used the same instrument to describe the actual groups they had participated in. Assuming that congruence between preferred and actual descriptions indicates a maximally attractive group, a measure of the groups' potential to satisfy the subjects' interests was computed from discrepancies on the three scales which each subject had called "most important" in making the group attractive to him (Satisfaction Potential).

In order to test the validity of the theoretical assumption that attraction would be a function both of the valence of membership and subjective probability of acceptance, correlations of each of the latter two variables with the prerejection Desire to Stay scores were obtained. Subjective Probability of Acceptance correlated .23 (biserial *r*, *ns*) and Satisfaction Potential .55 ($p < .01$) with Attraction to the Group. Scores obtained by multiplying the Subjective Probability and Satisfaction Potential variables correlated .59 ($p < .01$) with Desire to Stay. While combining the two variables thus seemed to add very little in predictive power to the Satisfaction Potential measure taken singly, inspection of Table 2 indicates that the addition of the Subjective Probability of Acceptance variable made a considerable difference when Satisfaction Potential scores were high. The validity of our assumption about the determinants of Attraction to the Group therefore appeared partially confirmed.

The before and after means for each of the three attraction questions in the four experimental conditions are listed in Table 3, together with the mean changes in attraction after rejection. The average level of attraction to the group was moderately high before rejection and did not differ significantly between experimental conditions. Two of the attraction measures, Desire to Stay and Frequency of Meeting, show a similar pattern of changes after rejection. On both measures, Strong Noninvidious rejection led to a decrease in attraction to the group, while neither of the Invidious rejection conditions produced any change.

The analysis of variance of these results, summarized in Table 4, indicates that the effect of reasons for rejection was significant at the 5% and 10% levels of confidence, respectively, for the Desire to Stay and Fre-

TABLE 2

MEAN DESIRE TO STAY BY SATISFACTION POTENTIAL AND SUBJECTIVE PROBABILITY OF ACCEPTANCE

($N = 57$)

Satisfaction potential	Subjective probability acceptance	Mean desire to stay	*N*	*SD*	*t*
High	High	8.71	17	1.12	8.20****
	Low	7.37	19	1.84	
Low	High	5.82	11	3.56	0.37
	Low	5.30	10	2.43	

**** $p < .001$.

TABLE 3

MEAN SCORES ON VARIABLES RELATED TO ATTRACTION BEFORE AND AFTER REJECTION

Variable	Rejection conditions			
	Mild non-invidious	Strong non-invidious	Mild invidious	Strong invidious
Desire to Stay				
Before	7.6	7.1	6.5	7.3
After	6.4	5.0	6.4	7.4
Change	−1.2	−2.1	−0.1	0.1
Frequency of Meeting				
Before	3.7	3.8	3.8	3.8
After	3.6	3.1	4.1	3.8
Change	−0.1	−0.7	0.3	0.0
Desire to Rejoin				
Before	6.7	6.9	6.3	8.0
After	6.7	5.7	6.4	6.8
Change	0.0	−1.2	0.1	−1.2
Desire for Contact				
After	4.2	3.8	4.6	3.3
Attraction to a New Group				
After	7.3	9.6	6.0	7.9

quency of Meeting measures.[2] The third measure of attraction, Desire to Rejoin, shows a different pattern of results, however. Reasons for rejection produced no effect at all on this measure, but degrees of rejection apparently produced differences sufficient to reach the 10% level of confidence. As Table 3 shows, there was no loss of attraction after mild rejection on this measure, but strong rejection led to some decrease in desire to rejoin the group. In summary, Hypothesis 1 was supported by data on changes in attraction to the group on two of our measures, but not on a third.

The differences in the results for the three measures unfortunately cannot be accounted for by lack of correlation between them. Desire to Rejoin correlated .74 ($p < .01$) with Desire to Stay and .65 ($p < .01$) with Frequency of Meeting; the latter two measures correlated .83 ($p < .01$) with each other ($N = 57$). Possibly the differences that appeared after the rejection treatment can be accounted for by the time reference of the questions. Desire to Stay and Frequency of Meeting refer to present intentions to continue to participate in the group, but Desire to Rejoin refers to future participation, which might be much less relevant to the subjects' present need for social reassurance.

According to Hypothesis 2 we would expect an interaction effect between reasons for and degrees of rejection. Although on the Desire to Stay and the Frequency of Meeting measures of attraction the differences in attraction loss are indeed greater for the two strong rejection conditions than for the mild rejection treatments, the predicted interaction effect does not occur. It appears therefore that Hypothesis 2 must be rejected on the basis of our data.

Two other results displayed in Table 3 remain to be discussed. Desire for Contact outside the experimental sessions was thought to reflect the subjects' liking for each other as persons rather than as members of the same bridge club. Although the means for the mild rejection treatments were slightly

TABLE 4

ANALYSIS OF VARIANCE SUMMARIES FOR VARIABLES RELATED TO ATTRACTION TO THE GROUP

Variable	Source	*MS*	*df*	*F*
Change in Desire to Stay	Degrees	1.69	1	
	Reasons	31.69	1	6.05**
	D × R	3.51	1	
	Error[a]	5.24	44	
Change in Frequency of Meeting	Degrees	2.52	1	
	Reasons	3.52	1	3.29*
	D × R	0.19	1	
	Error[a]	1.07	44	
Change in Desire to Rejoin[b]	Degrees	0.66	1	3.88*
	Reasons	0.01	1	
	D × R	0.00	1	
	Error[a]	0.17	44	
Desire for Contact	Degrees	80.20	1	2.99
	Reasons	00.40	1	
	D × R	26.70	1	
	Error[a]	26.82	44	
Attraction to a New Group	Degrees	52.10	1	7.54***
	Reasons	27.00	1	3.91*
	D × R	00.40	1	
	Error[a]	6.91	44	

[a] Error term based on pooled estimate.
[b] Scores adjusted by square root transformation to correct for heterogeneity of variance.
* $p < .10$.
** $p < .05$.
*** $p < .01$

[2] In order to rule out the possible effect of variations in attraction to the group before rejection, the data were also analyzed by the analysis of covariance method, controlling on prerejection attraction. The results of this analysis left the *F* values for the attraction measures essentially unchanged from those reported in Table 4.

TABLE 5

MEAN CHANGE IN PARTICIPATION RATES BY CONTENT AND EXPERIMENTAL CONDITIONS[a]

Content area	Rejection conditions			
	Mild noninvidious	Strong noninvidious	Mild invidious	Strong invidious
Task demanded	0.2	−0.3	−0.8	−1.3
Task related	−0.1	−0.1	−0.7	−0.8
Personal	−1.5	−1.8	−0.1	0.3
Total change	−1.4	−2.2	−1.6	−1.8

[a] Negative numbers indicate decreases in rates of participation.

higher than those for the strong rejection conditions, the difference between them is not statistically significant. The rejection treatments did not create any differences in desire to affiliate with the members of the bridge clubs, but this result can probably not be interpreted to mean that the rejection treatments produced no hostility towards members of the rejecting group. The experimenter was told of hostile feelings towards others in the group by a number of subjects during the post-experimental discussions.

Subjects were also asked to indicate their preference for joining another bridge club than the present experimental group. As the results for Attraction to Another Group show, strong rejection appeared to increase the subjects' willingness to join another group more than mild rejection in both invidious and noninvidious conditions ($p < .01$, Table 4). Reasons for rejection may also have had some effect, however; invidious rejection appears to depress the subjects' willingness to join another group slightly ($p < .10$, Table 4). It is of interest to note that the mean for attraction to another group in the Strong Noninvidious condition is greater than the mean for attraction to the experimental group *before* rejection. Apparently strong rejection tends to boost the subjects' interest in joining a bridge club to a higher absolute level.

Rates of verbal participation were calculated for the interaction period before and after rejection. Table 5 presents mean changes in rates of participation, broken down by content area. In general, participation decreases by about the same amount in all experimental conditions. The content area in which the decrease occurs, however, is quite different for the Invidious and Noninvidious conditions. Noninvidiously rejected subjects decrease their participation by reducing the number of *personal* remarks they make after rejection; their rates of participation in the *task demanded* and *task related* content areas remain virtually unchanged. The effect of reason for rejection on changes in personal content rates is significant at the 1% level of confidence (Table 6). Invidiously rejected subjects do not reduce personal remarks, as Table 5 shows, but do reduce the number of their remarks in the task demanded and task related categories; this latter effect falls short of statistical significance. If we interpret rate and content of participation as indicators of the subjects' motivation to participate in the group, it appears that noninvidious rejection leads to a withdrawal of personal interest in the group, although interest in the task remains unaffected. This kind of withdrawal fails to appear after invidious rejection, lending support to Hypothesis 1 about differences in tendencies to withdraw from the rejecting group between invidious and noninvidious rejection conditions.

DISCUSSION

It should be noted that the manipulation of degree of rejection by itself failed to reduce attraction to the group on two out of three measures, unlike the results of earlier experiments which showed a direct relation between lack of acceptance and resultant attraction to the group (Dittes, 1959; Dittes & Kelley, 1956; Kelley & Shapiro, 1954). The explanation may be found in differences between the earlier experiments and this one in the degree of commitment to a choice that

TABLE 6

ANALYSIS OF VARIANCE SUMMARY FOR CHANGES IN PERSONAL CONTENT OF COMMUNICATIONS

Source	*MS*	*df*	*F*
Degrees	0.02	1	
Reasons	38.52	1	9.85***
D × R	1.69	1	
Error[a]	3.88	44	

[a] Error term based on pooled estimate.
*** $p < .01$.

was implied in the attraction measurement procedures. Feather (1959) has pointed out the need to differentiate between measures of "attainment attractiveness, defined in terms of a wishful choice in a situation where there is no implied commitment," and "choice potential, defined by a choice in a situation where there is implied commitment." The procedures used in the present study definitely led subjects to believe they had to choose whether or not to attend future meetings of the group. As was seen above, the measure that is least like a direct choice because it refers to a hypothetical future event (Desire to Rejoin), produced results that at once differed from the other two attraction measures and were most like the results reported in earlier studies. In short, the results of this study conformed to the hypothesis about the effects of different reasons for rejection where measures of *choice* of participation in the group were involved (i.e., Desire to Stay and Frequency of Meeting). Desire to Rejoin the group in the future, which is more likely to measure *preference* for membership in the group without commitment to an actual choice, was not influenced by reason for rejection but only by the degree of rejection. If one takes the lack of Desire to Rejoin on the part of both strongly rejected groups to mean that membership in the group was actually unattractive to both of them, it can be concluded that the reason invidiously rejected subjects continued to choose membership in the rejecting group was probably that nonmembership had become more repulsive to them as a result of their invidious rejection.

It is likely that invidiously rejected subjects will not continue to choose membership in the rejecting group under all circumstances. In terms of the theoretical schema presented above, it is to be expected that over time (*a*) subjective probability of acceptance would decrease until the resultant attraction to the group approaches zero, and that (*b*) as the individual continues to experience rejection the valence of membership will in fact be lowered. In this connection it should be noted that the differences obtained in this study between invidious and noninvidious rejection cannot be explained in terms of differential expectations about being reaccepted into the group. The data do not show, for example, that noninvidious rejection led to a greater decrease in Subjective Probability of Acceptance than invidious rejection (Table 1). The evidence does support the view that reason for rejection is an important factor in determining the subjects' reactions to rejection. In particular, the more invidious the grounds for rejection, the more likely it is that new needs come into play that help determine the subjects' behavior. Under circumstances such as were employed in this study such needs may cause subjects to prefer efforts to join the rejecting group over leaving.

SUMMARY

A distinction was made between invidious rejection, defined as rejection for reasons that are likely to reflect upon the subjects' self-esteem, and noninvidious rejection. On the basis of the hypothesis that invidious rejection would arouse a need for social reassurance, it was predicted that (*a*) for any given degree of rejection, loss of attraction to the group will be smaller when the reason for rejection is invidious than when it is noninvidious, and (*b*) that the differences in loss of attraction between invidious and noninvidious conditions will be greater after strong than after mild rejection.

The predictions were tested in a 2×2 factorial experiment, employing two degrees of rejection and two different reasons for rejection, one invidious and the other noninvidious. Both attraction to the group and subjective probability of acceptance were measured before and after an experimentally induced rejection experience. The groups were experimentally created bridge clubs of four members, two of whom were confederates of the experimenter. There were 48 experimental subjects, all college students. The results with respect to attraction to the group generally supported Hypothesis 1, but not Hypothesis 2. In addition it was found that in the noninvidious condition subjects engaged in significantly less conversation of a personal nature after rejection. A secondary finding of interest was that willingness to join another group was greater after strong than after mild rejection, exceeding even the original level of attraction to the experimental group.

REFERENCES

BACK, K. W. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9–23.

DITTES, J. E. Attractiveness of group as function of self-esteem and acceptance by group. *J. abnorm. soc. Psychol.*, 1959, **59**, 77–82.

DITTES, J. E., & KELLEY, H. H. Effects of different conditions of acceptance on conformity to group norms. *J. soc. Psychol.*, 1956, **53**, 100–107.

FEATHER, N. T. Subjective probability and decision under uncertainty. *Psychol. Rev.*, 1959, **66**, 150–164.

JACKSON, J. M. A space for conceptualizing person-group relationships. *Hum. Relat.*, 1959, **12**, 3–16.

JACKSON, J. M., & SNOEK, J. D. Effect of invidious exclusion from a group on feelings toward self, others and on tendencies to conform. Paper presented at the sixty-seventh Annual Meetings of the American Psychological Association, Cincinnati, Ohio, 1959.

KELLEY, H. H., & SHAPIRO, M. An experiment on conformity to group norms where conformity is detrimental to group achievement. *Amer. sociol. Rev.*, 1954, **19**, 667–677.

LIBO, L. M. *Measuring group cohesiveness.* Ann Arbor: Research Center for Group Dynamics, 1953.

(Received January 20, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 64, No. 3, 183-191

SIMILARITY OF PERFORMANCE AS INFLUENCED BY INTERACTION, SELF-ESTEEM, AND BIRTH ORDER [1]

EZRA STOTLAND AND NICKOLAS B. COTTRELL

University of Washington

This study concerns the relationship between the amount of interaction between two persons in a group and the acquisition by one of the persons of an attribute of the other. It is hypothesized that the greater the interaction between the two, the more will one acquire an attribute of the other. In this fashion the effects of interaction per se can be studied, that is, the effects of interaction in which the information communicated by the interaction has no relevance to the outcome or dependent measure studied. This study also investigates the differential tendencies of group members of high and low self-esteem to acquire attributes in this manner.

The work of Michotte (1946) provides the theoretical source of this hypothesis. His studies of the perception of causal relationships between physical objects suggest that when one object is perceived as causing a change in a second object, the change in the second is perceived to belong to the first. In other words, a new unit or new entity is perceived, consisting of the first, causing object and of the change in the second object. This proposition has been extended theoretically to the area of social psychology by Heider (1958). He postulates that a person and the perceived consequence of his behavior will be seen as a unit.

The theory can be extended to the area of the perception of the interaction between two (or more) persons. By definition, when two people are interacting, each act by one of them is followed by, or causes, some event (e.g., change in behavior) in the other. The causing person and the event in the other person would then constitute a perceptual unit. For example, a drill sergeant barks a command at his squad of soliders and they respond immediately. The sergeant and the squad would be perceived by an observing officer, as well as by the soldiers and the sergeant, as constituting a single unit.

One additional assumption is needed to derive our hypothesis. The phenomenon of assimilation found in the study of the perception of physical objects is the tendency to perceive all parts of the same perceptual unit as having the same quality, e.g., degree of brightness. It is reasonable to expect that this assimilative tendency toward perception of the uniformity of parts of a unit would be greater in strong units than in weaker units. Therefore, it would be expected that the stronger the composite unit formed by the two persons interacting, the greater will be the tendency for the two persons to be perceived as having the same attributes. These similar attributes do not have to have anything directly to do with the quality, type, or nature of the interactions upon which the composite unit is based. In our example, the sergeant may be perceived to be an excellent marksman, so that his men come to raise their evaluation of their markmanship; the composite unit was based on his giving of orders on the drillfield, while the similar attribute, marksmanship, might be perceived only on the rifle range.

This relationship between interaction and perception of similarity of attributes has already been demonstrated for the case in which interaction leads one of the persons to perceive the other as being very similar to himself by attributing his own traits to the other (Stotland, Cottrell, & Laing, 1960). The present study concerns one person's acquiring an attribute he initially perceives in the other. The main hypothesis is that interaction between two persons will result in perceived similarity of their attributes.

The particular attribute (equivalent to marksmanship in the above example) em-

[1] This research was conducted under a grant from the National Institute of Mental Health (M 2433).

ployed to test this hypothesis in the present study is self-evaluation of performance on a task in a laboratory setting. The hypothesis applies to the individual's perception of his ability on the task. However, the effects of acquisition of this attribute may be deep enough to influence an individual's actual level of performance when he is given an opportunity to perform on the same task. A second hypothesis then predicts that when two persons have interacted to a high degree, the level of performance of one will directly influence the level of performance of the other more than when they have interacted to a low degree.

It is also assumed on the basis of a study by Stotland et al. (1960) that the boundaries of the self-conceptions of persons low in self-esteem are more penetrable than of persons high in self-esteem. In the present study, the penetration would be the person's perceiving himself to have the same level of ability on a task as a person with whom he has interacted. Thus, the third hypothesis of the present study is that persons with low self-esteem will behave more consistently with respect to the first two hypotheses than will persons high in self-esteem.

METHOD

Subjects. Students attending introductory and intermediate courses in psychology at the University of Washington during Fall Quarter, 1959, served as subjects. They were recruited at the beginning of the quarter to serve in a program of four experiments of which this research was one. The subjects received payment of $5.00 for serving in all four of the studies in the series.

The subjects were combined into six-person, same-sexed groups. An attempt was made to minimize prior acquaintance. Hence group members were chosen from different classes and within group similarities in last names and addresses were avoided.

On some occasions one or more of the scheduled subjects failed to appear at the scheduled time. At these times the experimenter summoned a paid assistant and ushered him in as if he were the tardy subject. These "stooges" had no special function. They followed instructions like the other subjects. The data gathered from these "stooges" were not considered in the analysis.

Procedure. When a group of six subjects arrived at the experimental room they were seated at a round table in the middle of which was a Lazy Susan with six compartments. The table was divided into six sections by fiberboard separators; thus, each subject could see only the tops of the other subjects' heads. On the table space in front of each subject were: an envelope containing 10 or 12 packs of eight syllables each, a 3″ × 5″ pad of pages, a box of paper clips, a pencil, a 3 × 5 × 3 wooden file box, and a record sheet. The subjects were told that they were first to form words from their packs of syllables. Each eight-syllable pack yielded four words. The subjects fastened pairs of fitting syllables together with paper clips. The words formed fell into one of six categories. These categories were animals, birds, cities of the United States, foreign countries, fruits and vegetables, and modes of transportation. The subjects' second job was to collect words of one category. The category assigned to them appeared on their file box. When they formed words in their own category, they put them in their file box. Some of the words they formed belonged in categories other than their own. These words they sent off to the other categories by means of the Lazy Susan. The Lazy Susan was divided into six sections and each section was labeled with one of the category headings. To send a word to another category the subject turned the Lazy Susan until he found the category he wanted to send to, then he put on the word and turned the Lazy Susan back until his own category again confronted him. Subjects were instructed to keep track on their record sheet of the number of words they sent to each other category and the number of words they received from each category. When the subjects sent a word they were instructed to attach a slip of paper with the name of their category written on it.

The final instruction to the subjects was that there was to be no talking during the study.

The only interaction the subjects had with each other was the sending and receiving of words. It is obvious that the distribution of words could be arranged so that varying patterns or amounts of word interchange would result. In this study the materials were arranged so that every person sent eight words to the animals category and received eight words from animals. All categories but animals formed 40 own words in addition, forming a total of 48; the animals category formed none of its own words, but did form 8 for each of the five other types of words, from a total of 40.

Each group worked for 20 minutes on the interaction task. Most subjects, but not all, finished all their word assembling. After this they were told they were to try a new task. First, the experimenter had one of the group members demonstrate the new task. In the Interaction condition the demonstrator or model was the person who had collected the animal words and hence had been the center of the interaction (interchange of names). In the No Interaction condition the demonstrator was a person from the periphery of the prearranged interaction network; no one but the collector of animal words had exchanged words with the demonstrator in this condition. Groups were alternately assigned to these conditions.

The second task consisted of listening to a tape recorded reading of a passage in a foreign language (the language was Hindi) and keeping a mental count of the number of times a certain sound occurred. There were two passages and the person's score was the sum of the errors on each passage.

The demonstrators were first given the tape task; when it was over, the demonstrator wrote down his count. The demonstrator listened to the tape over a set of earphones so that the other subjects could not hear the tape while he was listening to it. In half the cases the demonstrator was told in the presence of the group that he had been very accurate in his count, his count being only 15% away from perfect. This accuracy, the experimenter said, was proof of his paying close attention.

This condition was called the Good Evaluation of Model's Performance. In the other half of the sessions, the demonstrator was told that he had been rather inaccurate in his counting, his count being 85% away from perfect. This inaccuracy, the experimenter said, was probably due to not paying attention. This condition is called the Poor Evaluation of the Model's Performance. Groups were assigned alternately to these two conditions.

The other subjects were then asked to rate how well they expected to do on the task. Then the subjects were given their own sets of earphones and listened to the two passages, recording their mental count at the end of each selection. They then rated their own performances. The subjects then indicated on a questionnaire how well they had known each of the other subjects, their level of interest in the listening task, and attractions to the group qua group, and desire to know each of the other group members. All of these ratings were made on seven-point scales.

The reasons that this task was selected were, first, that it was completely unfamiliar to the subjects and they would have no standards available upon which to base expectations of own performance; second, the model's performance would be judged only by the experimenter; third, the subjects could all perform the task simultaneously without knowledge of each other's level of performance; fourth, previous research using this task indicates that performance on this task is susceptible to social influence.

The data from subjects who were the center of the interaction network or were demonstrators were discarded from the analysis. At a later session, the study was explained to the subjects; also, their prior training in Hindi was ascertained. Any subjects who had prior experience with Hindi were dropped from the sample. Any subjects in the No Interaction condition who by mistake interacted with the demonstrator were dropped from the sample. Subjects who indicated they had known their demonstrator before serving in the study were also dropped from the sample. The experimental sessions lasted about an hour.

During a previous session of the four in which the subjects participated, they took a modification of the *Q* sort to measure self-esteem. This measure is described in Stotland et al. (1960).

RESULTS AND DISCUSSION

Total sample. The general hypothesis of this study is that the Interaction condition would lead the subjects to acquire an attribute of the demonstrator. This acquirable attribute was paying attention or not paying attention to the tape task (that is, good or poor model evaluation). The first hypothesis leads to the prediction that in the Interaction condition, the subjects will predict that they will do better on the tape task when the demonstrator ostensibly performs well than when he ostensibly performs poorly. The mean predictions of performance are presented in Table 1. The data show that the hypothesis is confirmed with respect to this prediction. The first hy-

TABLE 1

PERFORMANCE AND RATINGS[a] IN EXPERIMENTAL CONDITIONS

Evaluation of model's performance	Interaction condition		No interaction condition		t
	Good a	Poor b	Good c	Poor d	
1. Prediction of performance	3.04	3.55	3.45	3.45	a vs. b: 2.309**
					a vs. c: 1.948*
2. Number of errors	5.00	8.00	6.15	7.25	a vs. b: 2.039**
3. Evaluation of performance	2.33	2.65	2.75	2.90	
4. Desire to know others (not demonstrator)	14.93	14.07	16.05	15.63	
5. Desire to know demonstrator	3.59	3.24	3.85	3.84	
6. Attraction to group	3.48	3.41	3.60	3.50	
7. Interest in tape task	2.96	2.83	3.30	2.79	
N	27	29	20	19	

[a] The higher the rating, the less the quantity indicated (except for Row 2).
* $p < .06$, two-tailed.
** $p < .05$, two-tailed.

TABLE 2

PERFORMANCE AND RATINGS[a] IN EXPERIMENTAL CONDITIONS FOR SUBJECTS HIGH AND LOW IN DESIRE TO KNOW THE DEMONSTRATOR

Evaluation of model's performance	Interaction condition		No interaction condition		
	Good a	Poor b	Good c	Poor d	
Subjects High in Desire to Know Demonstrator					
1. Prediction of performance	3.00	3.69	3.78	3.67	a vs. b: 2.060*
2. Number of errors	5.93	7.25	5.38	7.56	
3. Evaluation of performance	2.40	2.62	2.62	2.89	
4. Desire to know others (not demonstrator)	12.33	11.63	14.75	11.78	
5. Attraction to group	3.27	2.94	3.38	3.11	
6. Interest in tape task	2.73	2.69	2.88	2.22	
N	15	16	8	9	
Subjects Low in Desire to Know Demonstrator					
1. Prediction of performance	3.08	3.39	3.50	3.30	
2. Number of errors	3.83	8.92	6.67	7.30	a vs. b: 2.161*
3. Evaluation of performance	2.25	2.69	2.83	2.90	
4. Desire to know others (not demonstrator)	10.17	17.08	16.92	19.10	c vs. d: 2.103*
5. Attraction to group	3.75	4.00	3.75	3.90	
6. Interest in tape task	3.17	3.50	3.58	3.56	
N	12	13	12	10	

[a] The higher the rating, the less the quality rated (except for Row 2).
* $p < .05$, two-tailed.

pothesis also leads to the prediction that the subjects will evaluate their performance (after working on the task) better when the model does better in the Interaction condition. The prediction is not supported by the data in Table 1. The first hypothesis also suggests the corollary that in the No Interaction condition, neither of the predictions made above will be supported. The data of Table 1 support this corollary. However, the most stringent confirmation of this first hypothesis would be to find that the difference between the good and poor Interaction conditions is greater than the difference between the good and poor No Interaction conditions. The significance of the difference was tested by a method suggested by Walker and Lev (1953, p. 158) which is essentially a *t* test of the difference between differences between means. This test was not significant for prediction of performance, or evaluation of performance. In general, however, the results provide some support for this hypothesis with respect to two of the predictions made from it.

The second hypothesis was that in the Interaction condition, the subjects will actually perform better on the listening task when the model had done well than when he had done poorly. Table 1 shows that this prediction is supported by the number of errors made in each condition. As would be expected from the hypothesis, there is no significant difference between the two No Interaction conditions. However, the Walker-Lev test of the difference between differences between means was not significant. Thus, the study provides some support for the hypothesis that the social influence in the present study influences the subjects' performances on the task.

Also presented in Table 1 are the mean ratings of attraction to the other members of the group, to the demonstrator, and to the group as a whole, and interest in the listening task, all made after the subjects had performed the task. None of the differences among the conditions is significant, nor are the Walker-Lev tests among the means, showing that the experimental conditions tend to influence performance variables significantly without influencing attraction variables.

Attraction. The work of Back (1951) and others has suggested that high attraction is

associated with conformity. Subjects who responded to the experimental conditions with high attraction to the demonstrator (model) may therefore have reacted differently from those who responded with low attraction. To test this possibility, the subjects were divided at the median of attraction to the demonstrator, and the experimental differences examined. In Table 2, it can be seen that attraction is not necessarily associated with the subject's conforming to the level of performance of demonstrator. Among subjects highly attracted, only the subjects' predictions of their performances showed the expected condition difference significantly. Subjects of low attraction to the model, on the other hand, showed the expected condition differences in their actual performances, making considerably fewer errors in the Interaction condition when the model had done well than when he had done poorly. It is quite clear, then, that the processes examined in the present study are different from the processes involved in studies when attraction does lead to conformity. With regard to both these variables, however, Walker-Lev tests of the difference between differences do not yield significant t ratios. Therefore, these conclusions must be regarded tentatively.

Prior acquaintance. The experimental results would reasonably be expected to have been influenced by the amount of prior acquaintance among the subjects. Table 3 shows the results separately for those subjects with some and with no prior acquaintance with other subjects. The experimental hypotheses are supported primarily by subjects with no prior acquaintance with other subjects. These subjects in the Interaction condition predict better performances, actually do better, and evaluate their performances better when the demonstrator performs better

TABLE 3

PERFORMANCE AND RATINGS[a] IN EXPERIMENTAL CONDITIONS FOR SUBJECTS WITH SOME AND NO PRIOR ACQUAINTANCE WITH OTHER SUBJECTS

Evaluation of model's performance	Interaction condition		No interaction condition		t
	Good a	Poor b	Good c	Poor d	
Subjects with No Prior Acquaintance					
1. Prediction of performance	2.81	3.45	3.25	3.64	a vs. b: 2.262**
2. Number of errors	3.69	8.65	8.33	5.86	a vs. b: 2.998*** a vs. c: 2.134**
3. Evaluation of performance	2.12	2.80	2.67	2.64	a vs. b: 2.344** a vs. c: 1.887*
4. Desire to know others (not demonstrator)	15.75	14.90	15.83	15.85	
5. Desire to know demonstrator	3.50	3.25	3.83	4.00	
6. Attraction to group	3.38	3.85	3.50	3.36	
7. Interest in tape task	2.56	2.85	3.75	2.92	a vs. c: 2.926***
N	16	20	12	14	
Subjects with Some Prior Acquaintance					
1. Prediction of performance	3.36	3.78	3.75	3.00	
2. Number of errors	6.91	6.55	2.88	10.50	a vs. c: 2.051* c vs. d: 1.906*
3. Evaluation of performance	2.64	2.33	2.87	3.50	b vs. d: 2.446**
4. Desire to know others (not demonstrator)	13.73	13.33	16.38	15.17	
5. Desire to know demonstrator	3.73	3.22	3.88	3.50	
6. Attraction to group	3.64	2.44	3.75	3.83	a vs. b: 2.351** b vs. d: 2.572**
7. Interest in tape task	3.45	2.78	2.62	2.50	
N	11	9	8	6	

[a] The higher the rating, the less the quality rated (except for Row 2).
* $p < .10$, two-tailed.
** $p < .05$, two-tailed.
*** $p < .01$, two-tailed.

TABLE 4

PERFORMANCE AND RATINGS[a] IN INTERACTION CONDITIONS FOR SUBJECTS HIGH AND LOW IN NUMBER OF NAMES SENT AND RECEIVED

Evaluation of model's performance	High number		Low number		t
	Good a	Poor b	Good c	Poor d	
1. Prediction of performance	3.68	3.62	3.00	3.50	
2. Number of errors	5.39	6.46	4.64	9.25	c vs. d: 2.275*
3. Evaluation of performance	2.15	2.46	2.50	2.81	
4. Desire to know others (not demonstrator)	14.69	14.85	15.14	13.44	
5. Desire to know demonstrator	3.77	3.15	3.43	3.31	
6. Attraction to group	3.77	3.15	3.21	3.62	
7. Interest in tape task	2.92	2.54	2.93	3.06	
N	13	13	14	16	

[a] The higher the rating, the less the quality rated (except for Row 2).
* $p < .05$, two-tailed.

than when he performs worse. In fact, these effects are so strong that subjects in the Interaction condition with a good model perform better and evaluate themselves better than those in the No Interaction condition with a good model. On the other hand, subjects who have at least some prior acquaintance with the other subjects behave differently. They do not manifest to a significant degree any of the differences predicted between the good and poor Interaction conditions. In fact, a Walker-Lev test shows that the subjects of no prior acquaintance show these differences to a greater degree. For the number of errors, the Walker-Lev t ratio is 1.738 ($p < .10$, one-tailed), while it is 1.740 for evaluation of performance ($p < .10$, one-tailed).

Futhermore, in the No Interaction conditions, the subjects of some prior acquaintance react in a completely unexpected manner. In these conditions the subjects perform better when the model does better; and the mean number of errors is significantly less in the No Interaction condition with a good demonstrator than in the Interaction condition with a good one. Moreover, a Walker-Lev test yields a t ratio of 2.655 ($p < .02$, two-tailed), showing that the subjects of some prior acquaintance in the No Interaction condition do react differently from those of none. The authors confess that, without building a pseudological structure of tenuous assumptions and derivations, they are at a loss to explain the behavior of the persons who had some prior acquaintance.

Number of names sent and received. In the actual administration of the Interaction condition, there were two factors preventing all of the subjects from exchanging as many names with the to-be demonstrator as the procedure called for. These factors were: slowness of assembling names by particular demonstrators and subjects; errors by the demonstrator or subjects in sending names. Although these errors were minimal, the subjects in the Interaction condition could still be divided into two groups according to the number of names sent and received. It would be expected that those sending and receiving a high number of names would confirm the hypotheses better than those sending and receiving a low number. The results shown in Table 4 do not support this expectation. With respect to actual performance, the subjects with a low number of names sent and received behaved in line with the model's performance, while the higher did not. However, a Walker-Lev test does not demonstrate a significant difference between the two groups of subjects.

Nevertheless, the surprising direction of the difference does require discussion. A possible explanation is that the subjects with a low number of exchanges of syllables with the demonstrator were still attempting to achieve a certain quality of relationship with him, which would be achieved by making oneself similar to him. On the other hand, those with a high number of exchanges had achieved a satisfactory level of exchange with him and,

therefore, did not have to behave like him to achieve a certain quality of relationship.

Self-esteem. The third hypothesis predicted that subjects low in self-esteem would be more responsive to the experimental differences in the expected ways than would subjects high in self-esteem. The results shown in Table 5 indicate that exactly the opposite tended to occur. The subjects with low self-esteem evidenced no significant differences at all between experimental conditions on any dependent variable. On the other hand, in the Interaction condition, subjects high in self-esteem predicted that they would perform better, actually performed better, and evaluated their performance better when the demonstrator did well than when he did poorly. No differences occurred for these subjects in the No Interaction condition. In fact, the subjects in the Interaction condition with a good demonstrator predicted better performance than those in the No Interaction condition with the same quality of demonstrator. A Walker-Lev test of the significance of the differences in the two Interaction conditions between the subjects high and low on self-esteem yields a ratio of 1.913 ($p < .06$, two-tailed) for evaluation of performance. Parallel tests for prediction of performance and number of errors do not yield significant t ratios.

The obtained pattern of results shows that the overall result reported in Table 1 for the whole sample was generated more by subjects high in self-esteem than by those lower in self-esteem. These results are that the subjects perform better in the Interaction condition with a good demonstrator. This may have been produced by the high self-esteem subjects' reacting to the experimental conditions by high self-evaluation. However, without the appropriate social situation, the high self-esteem persons do not behave in this fashion, since they did not show any difference between the two No Interaction conditions. This tendency of persons with high self-esteem to react to social situations

TABLE 5

PERFORMANCE AND RATINGS[a] IN EXPERIMENTAL CONDITIONS FOR SUBJECTS HIGH AND LOW IN SELF-ESTEEM

Evaluation of model's performance	Interaction condition		No interaction condition		t
	Good a	Poor b	Good c	Poor d	
Subjects High in Self-Esteem					
1. Prediction of performance	2.80	3.64	3.14	3.62	a vs. b: 2.509**
2. Number of errors	3.33	7.77	7.86	7.54	a vs. b: 4.976***
3. Evaluation of performance	1.87	2.64	2.71	2.85	a vs. b: 2.587**
					a vs. c: 2.211*
4. Desire to know others (not demonstrator)	15.27	13.21	15.86	15.25	
5. Desire to know demonstrator	3.60	3.21	3.86	3.75	
6. Attraction to group	3.40	3.57	3.14	3.89	
7. Interest in tape task	2.53	2.43	3.14	2.77	
N	15	14	7	13	
Subjects Low in Self-Esteem					
1. Prediction of performance	3.30	3.47	3.67	3.14	
2. Number of errors	6.50	8.27	5.50	6.71	
3. Evaluation of performance	2.80	2.67	2.83	3.00	
4. Desire to know others (not demonstrator)	14.00	14.87	16.83	16.29	
5. Desire to know demonstrator	3.30	3.27	4.00	4.00	
6. Attraction to group	3.60	3.27	3.83	3.71	
7. Interest in tape task	3.20	3.70	3.42	2.83	
N	10	15	12	7	

[a] The higher the rating, the less the quality rated (except for Row 2).

* $p < .05$, two-tailed.
** $p < .02$, two-tailed.
*** $p < .001$, two-tailed.

TABLE 6

PERFORMANCE AND RATINGS[a] IN EXPERIMENTAL CONDITIONS FOR FIRST BORN AND ONLY CHILDREN AND FOR LATER BORNS

Evaluation of model's performance	Interaction condition		No interaction condition		t	p[b]
	Good a	Poor b	Good c	Poor d		
First and Only Children						
1. Prediction of performance	3.20	3.64	3.29	3.31		
2. Number of errors	5.67	5.65	5.00	8.62		
3. Evaluation of performance	2.33	2.54	2.43	3.08	c vs. d: 2.09	<.05
4. Desire to know others (not demonstrator)	14.73	12.43	15.71	16.08		
5. Desire to know demonstrator	3.93	3.14	3.86	4.00		
6. Attraction to group	3.33	2.86	3.43	3.00		
7. Interest in tape task	3.47	3.14	3.64	3.62		
N	15	14	14	12		
Later Borns						
1. Prediction of performance	2.83	3.46	3.80	3.67	a vs. b: 1.839 a vs. c: 2.173	<.10 <.05
2. Number of errors	4.17	10.85	9.20	5.17	a vs. b: 2.819 b vs. d: 1.988	<.02 <.07
3. Evaluation of performance	2.33	2.85	3.40	2.50	a vs. c: 2.246 c vs. d: 1.965	<.05 <.08
4. Desire to know others (not demonstrator)	15.17	15.69	16.40	14.00		
5. Desire to know demonstrator	3.17	3.31	3.60	3.33		
6. Attraction to group	3.50	3.92	3.60	3.17		
7. Interest in tape task	2.42	2.92	3.00	2.67		
N	12	13	5	6		

[a] The higher the rating, the less the quality indicated (except for Row 2).
[b] All p values two-tailed.

mainly when the consequences of the reaction would be ego enhancing has also been found by Burnstein, Stotland, and Zander (1961), and Stotland, Thorley, Thomas, Cohen, and Zander (1957), and has been discussed by Cohen (1959). These persons evidently maintain their high self-esteem through mechanisms similar to the one manifested here. Persons low in self-esteem may react more readily in social situations in which social pressure is exerted in the form of an explicit request, demand, or expectation by other persons. This was the case in Hovland and Janis (1959) and Stotland et al. (1957). In the present study, no requests, demands, or expectations were explicitly communicated to the subjects regarding their behaving like the demonstrator. Therefore, the low self-esteem persons behaved the same in both the Interaction and the No Interaction conditions.

Birth order. Schachter (1959) has demonstrated the significance of birth order for reactions to social situations. To explore this variable further, the subjects were divided into two groups, first and only children, and later born children, and the results tabulated separately for each. Table 6 clearly shows that the later borns supported the experimental hypothesis much more than the first born and only children. The first born and only children show only one significant difference between conditions, which was in their evaluation of performance in the No Interaction conditions. On the other hand, in the Interaction condition, the later borns predicted better performance and actually performed better when the model was good than when he was poor. In the interaction conditions, the Walker-Lev test between the two samples yields a t ratio of 2.278 ($p < .05$) for number of errors, showing that the later borns reacted more to the experimental differences. In

fact, in the No Interaction condition there was some tendency for the later borns to react in a manner completely opposite to the way they reacted in the Interaction condition. That is, they evaluated their performance better when the model was worse. Further indications of the reversal are that, when the model was good they made better predictions and evaluations in the Interaction condition than in the No Interaction condition; and that, when the model was poor, they performed worse in the Interaction condition than in the No Interaction condition. The Walker-Lev test, between the two samples in the No Interaction conditions, produces a t ratio of 1.779 ($p < .10$, two-tailed) for number of errors, and 2.719 ($p < .02$, two-tailed) for evaluation of performance.

SUMMARY

In the present study a condition of high interaction was created in one task. The effects of this interaction on perceived similarity to the high interactor and to his level of performance were tested in another task. The intention was to test for interaction per se without the involvement of the content of the interaction. Twenty groups of six subjects each (all college students) were formed to carry out a word assembly task so arranged that one of the six would be the center of the interaction. Subsequently, the groups were given the task of detecting the number of repetitions of the same sounds in a tape recording in a foreign language. This second part of the experiment was carried out by assigning groups to one of four situations: the high interactor from the first task was the model with public praise for his performance; the high interactor was the model with public criticism of his performance; a subject not central to the interaction in the first task was the model with public praise for performance; and a subject not central to the interaction in the first task was the model with public criticism of his performance. Measures were taken of the subjects' predictions of the performance, of their actual performance, and their evaluation of it.

The findings were:

1. Consistent with the hypothesis of the study, the subjects rated their performances, and actually performed on the counting task, more in line with the interaction model's ostensible level of performance than with the no interaction model's.

2. Subjects low in attraction to the model appeared to behave more consistently with the hypothesis in their actual performances than did subjects high in attraction.

3. Subjects who had no prior acquaintance with other subjects confirmed the hypothesis more than those of some prior acquaintance.

4. Subjects who for a variety of reasons exchanged a low number of words with the interaction model performed on the counting tasks more consistently with his ostensible level of performance than did those who exchanged a high number.

5. Subjects high in self-esteem behaved more consistently with the hypothesis above than did subjects low in self-esteem.

6. Subjects who had been later born children behaved more consistently with the experimental hypothesis than did the first and only children.

REFERENCES

BACK, K. W. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9–23.

BURNSTEIN, E., STOTLAND, E., & ZANDER, A. Similarity to a model and self-evaluation. *J. abnorm. soc. Psychol.*, 1961, **62**, 257-264.

COHEN, A. R. Some implications of self-esteem for social influence. In C. Hovland & I. Janis (Eds.), *Personality and persuasibility*. New Haven: Yale Univer. Press, 1959. Pp. 102–120.

HEIDER, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.

HOVLAND, C., & JANIS, I. (Eds.) *Personality and persuasibility*. New Haven: Yale Univer. Press, 1959.

MICHOTTE, A. E. *La perception de la causalite*. Paris: Vocin, 1946.

SCHACHTER, S. *The psychology of affiliation*. Stanford, Calif.: Stanford Univer. Press, 1959.

STOTLAND, E., COTTRELL, N. B., & LAING, G. Group interaction and perceived similarity of members. *J. abnorm. soc. Psychol.*, 1960, **61**, 335–340.

STOTLAND, E., THORLEY, S., THOMAS, E. J., COHEN, A. R., & ZANDER, A. The effects of group expectations and self-esteem upon self-evaluation. *J. abnorm. soc. Psychol.*, 1957, **54**, 55–63.

WALKER, HELEN M., & LEV, J. *Statistical inference*. New York: Holt, 1953.

(Received January 23, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 64, No. 3, 192-205

DIMENSIONS OF SUBJECTIVE ADJUSTMENT[1]

JOSEPH VEROFF, SHEILA FELD,[2] AND GERALD GURIN

Survey Research Center, University of Michigan

Research programs assessing mental health have a broad range of problems that are possible of exploration. The criteria used to evaluate health or illness are diverse. Problems of social adjustment, aspects of self-esteem, psychosomatic complications, creativity all can be investigated. Or one can simply identify disturbance pragmatically by isolating those admitted to mental hospitals or those who have been exposed to some sort of psychiatric treatment. Recent thinking about this confusing diversity of criteria for evaluating mental health and illness has converged on the reasonable viewpoint that different criteria for such evaluations are not only equally possible and equally valid but are required for a complete appraisal of mental functioning (Eaton, 1951; Fiedler, Dodge, Jones, & Hutchins, 1958; Jahoda, 1950, 1953, 1958; Scott, 1958; Smith, 1950, 1959).

In the present study, a factorial classification of various self-descriptions of adjustment and distress, the data for inferences about mental health were limited to self-reports, thereby ruling out considerations of criteria for mental health that are based on other sources of inference, such as the assessment of actual behaviors or the diagnoses of an "objective" clinical observer or the judgments of peers. But even within this focus on self-evaluations of psychic distress, there exists the possibility of using several different kinds of dimensions for evaluating "good" and "poor" mental health. Self-descriptions are often assumed to be a single criterion to be compared with criteria based on evaluations from psychiatrists, friends, etc. (Eaton, 1951; Fielder et al., 1958). But there are many facets to self-descriptions. And many ways of defining felt distress appear in the research literature on mental health. Different indices of dysfunctioning focus on different aspects of psychic adjustments: psychosomatic complaints (MacMillan, 1957), ways of perceiving the self (Rogers, 1951), life satisfactions in marriage (Terman & Wallin, 1949), in work (Herzberg, Mausner, & Snyderman, 1959; Morse, 1953), life happiness in general (Inkeles, 1960). Each of these approaches is by itself inadequate for fully identifying distress. Usually, the investigator concentrates on one aspect of these subjective evaluations. In so doing he may be trapped into identifying this aspect of mental health as the most important criterion for defining mental health. In the present study, rather than choosing among these various criteria of distress, a multiple criterion approach was adopted. The research to be reported suggests that multiple criteria are needed in a description of subjective adjustment, and further suggests what these criteria might be psychologically.

The present paper is a further report on a nationwide sample survey recently reported on by the present authors (Gurin, Veroff, & Feld, 1960). A total of 2,460 adult respondents were selected by area sampling probability methods to constitute a representative cross-section of adults, 21 years of age or older, living in private households in the United States (Kish, 1953). This sample omits the hospitalized population. They were interviewed about their lives, their experiences of distress or satisfaction in many *different* kinds of ways. In utilizing a multiple criterion approach to the measurement of feelings of adjustment, each index was considered separately, on its own terms. No one measure was weighted more heavily than the other as a "better" measure of psychic distress. Rather, we assumed that unhappiness, the sense of inadequacy, tension in a marriage or with children, problems at work, constant worries,

[1] The investigations reported in this paper are based on data obtained within a national sample survey supported by the Joint Commission on Mental Illness and Health. Analysis of these data was supported by a grant from the National Institute of Mental Health, United States Public Health Service, Project M2280.

[2] Now at Yale University.

feelings of an impending nervous breakdown, all reflect different areas of potential distress. In the analyses already reported, the information in the survey was examined to see how *each* index varied from social group to group, and how *each* was related to patterns of seeking professional help for personal problems.

A major question remains from these analyses: How are the various indices covering different aspects of life adjustments related one to another—especially for groups for whom all indices were relevant—married women with children and employed married men with children who fall into our non-institutionalized United States population? This is the problem for this paper. In this study, factor analytic techniques were used to assess the communalities in response to the diverse questions asked of the national sample. The emerging factor structure may be indicative of basic psychological factors underlying the determination of self-descriptions of adjustment, and as such it represents an empirical basis for theorizing about the significance of these disparate measures.

The factor analytic technique allows the researcher to look at clusters of responses in different ways. It may very well be in our analysis that certain clusters will reflect accommodations to strong cultural pressures to perceive the self in a prescribed manner. Or it may be that the clusters reflect deeply felt reactions to the common psychological problems faced by most people in American society. In general we will be following the latter interpretation of the significance of the analyses. We recognize that the other interpretation is possible and will highlight the alternate view where it appears appropriate.

We have approached the problem of investigating self-descriptions with the assumption that separate components to self-evaluations were more than likely. Orthogonal factors generating intercorrelations of responses are part of the assumptive framework used in procedures for this study in factor extraction and rotation. This assumption will not guarantee multiple dimensions of self-descriptions, but it favors such a possibility of data analysis.

All the indices of adjustment are based on self-descriptions. As such they raise the question of honesty of response. Some distortion in self-report is inevitable in a study of subjective adjustment. However, we would expect such distortion to be mainly reflected in a generalized tendency to admit or deny distress in the different life-areas investigated in this study, rather than in systematic distortions on certain of the measures. Since a factor analytic study would be mainly affected by systematic rather than generalized distortions, we might expect the error springing from such distortions to be less critical for the analyses reported here than in other analyses—for example, those which would involve interpretations of the absolute values of the responses to the different questions on which the adjustment measures are based. However, some systematic distortion may exist, and as such would add a source of error to the results.

Error springing from the problem of honesty of response may also be introduced when factor structures are compared for different subgroups of the population. For example, in the comparison between men and women that will be discussed in this paper, the question may arise as to whether the sex differences were affected by differences in either the prescribed stereotypes of responding for men or for women or in how willing men and women are to talk to an interviewer (usually a woman) in an interview situation. However, as will be noted in the later discussions, the similarities in the factor structures of men and women are more striking than the differences, and the differences that do exist do not seem relevant to the question of differences in honesty of response.

In the present study we have considered only a portion of the total sample of 2,460 subjects. Three interview forms were used in the study, each assigned randomly to one-third of the subjects. One-third of the total sample of men and two-thirds of the sample of women were asked all the questions dealt with in the present paper; certain questions were applicable only to currently married subjects with children or, among the men, to working subjects. The final *N*s are 542 women and 255 employed men who are currently married and have children.

PROCEDURE

The expression of distress was measured in six areas: general feelings of distress, attitudes towards the self, marital adjustment, adjustment as a parent, job adjustment (men only), and psychological and psychosomatic symptom complaints. The separate indices are listed in Table 1 under these headings together with the questions on which each index

TABLE 1

INDICES USED IN FACTOR ANALYSES OF SELF-DESCRIPTIONS OF DISTRESS

Questions	Indices
General Feelings of Distress	
Everybody has some things he worries about more or less. What kinds of things do you worry about most?	Worrying
Do you worry about such things a lot, or not very much?	All the time Sometimes Never
Taking things all together, how would you say things are these days—would you say you're *very happy*, *pretty happy*, or *not too happy* these days?	Unhappiness Not too happy Pretty happy Very happy
Compared to your life today, how do you think things will be 5 or 10 years from now—do you think things will be happier for you than they are now, not quite as happy, or what?	Future unhappiness Not as happy About the same Happier
Have you ever felt that you were going to have a nervous breakdown?	Nervous breakdown Yes No
Attitudes toward Self	
People are the same in many ways, but no two people are exactly alike. What are some of the ways in which you're different from most other people?	Lack of uniqueness of self Sees no differences from others Mentions differences
Same as above.	Lack of self-acceptance (degree to which subject gives negative evaluation of the self in describing differences from others) Negative Ambivalent Neutral Positive Very positive
Many people when they think about their children, would like them to be different from themselves in some ways. If you had a son (daughter) how would you like him (her) to be different from you?	Shortcomings in the self Wants child to be different Does not want child to be different
How about your good points? What would you say were your strongest points?	Lack of strong points in the self Sees no strong points Mentions strong points
Martial Adjustment	
Many men (women) feel that they're not as good husbands (wives) as they would like to be. Have you ever felt this way?	Feelings of marital inadequacy
What kinds of things make you feel this way?	
Do you feel this way a lot of times, or only once in a while?	A lot of times Once in a while Never
Taking things all together, how would you describe your marriage—would you say your marriage was very happy, a little happier than average, just about average, or not too happy?	Marital unhappiness Not too happy Average A little happier than average Very happy
Even in cases where married people are happy there have often been times in the past when they weren't too happy—when they had problems getting along with each other. Has this been true for you?	Experience of problems in marriage Mentions problems Never had problems
Adjustment as a Parent	
First, thinking about a man's (woman's) life, how is a man's (woman's) life changed by having children?	Negative orientation to children: Coder rating of change seen as accompanying children Negative Neutral Positive
Most parents have had some problems in raising their children. What are the main problems you've had in raising your children?	Experience of problems in raising children Mentions problems Never had problems
Many men (women) feel that they're not as good fathers (mothers) as they would like to be. Have you ever felt this way?	Feelings of inadequacy as a parent
(If yes) What kinds of things have made you feel this way? Have you felt this way a lot of times, or only once in a while?	A lot of times Once in a while Never
Psychological and Psychosomatic Symptom Complaints	
Do you ever have any trouble getting to sleep or staying asleep? Check one: Nearly all the time (1)[b]; Pretty often (2); Not very much (3); Never (4).	Symptom Factor 1: Psychological Anxiety[a]
Have you ever been bothered by nervousness, feeling fidgety and tense? Check one: as above.	Six ranks ranging from very high to very low scores
Do you feel you are bothered by all sorts of pains and ailments in different parts of your body? Yes (2); No (4).	Symptom Factor 2: Physical Ill Health[a]
For the most part, do you feel healthy enough to carry out the things you would like to do? Yes (4); No (2).	Three ranks
Do you find it difficult to get up in the morning? Check one: Nearly all the time (1); Pretty often (2); Not very much (3); Never (4).	Symptom Factor 3: Immobilization[a]
Are you troubled by your hands sweating so that you feel damp and clammy? Check one: Many times (1); Sometimes (2); Hardly ever (3); Never (4).	Six ranks

TABLE 1—Continued

Questions	Indices
Have you every been bothered by shortness of breath when you were not exercising or working hard? Check one: Many times (1); Sometimes (2); Hardly ever (3); Never (4).	Symptom Factor 4: Physical Anxiety[a] Six ranks
Have you ever been bothered by your heart beating hard? Check one: as above. Summation of 16 symptom questions used in Stirling County study (MacMillan, 1957).	Summary Symptom Score [c] Ten ranks
Job Adjustment	
Taking into consideration all the things about your job, how satisfied or dissatisfied are you with it?	Job dissatisfaction Dissatisfied Ambivalent Satisfied Very satisfied
Have you ever had any problems with your work—times when you couldn't work or weren't getting along on the job, or didn't know what kind of work you wanted to do?	Experience of problems with job Mentions problems Never had problems
How good would you say you are at doing this kind of work—would you say you were *very good, a little better than average, just average,* or *not very good?*	Feelings of inadequacy on job Not very good Average Little better than average Very good

[a] These symptom factor indices are based on prior factor analyses of a symptom list of 20 items which were taken, with slight changes, from the Health Opinion Survey (MacMillan, 1957). Each index represents the two items most clearly typifying the factor; these items are ones having not only high loadings on the factors they are assumed to represent, but also minimal loadings on other factors.
[b] Scores on the symptom factors are the sum of the values in parentheses.
[c] For a list of these items see Gurin, Veroff, and Feld (1960, Table 7).

was based. For a full description of the coding procedures for each index and the complete interview schedule the reader is referred to *Americans View Their Mental Health* (Gurin et al., 1960).

For purposes of intercorrelation, the responses to the questions constituting each index were ranked—with a rank of 1 indicative of highest distress. As can be seen in Table 1, the number of rank positions varied from index to index, with seven of the indices having only two rank positions.

Separate matrices of intercorrelations of these indices were obtained for men and women. Since the sex differences in response to some of the separate indices were very great, it was felt that the posssibility of distinctive factor structures emerging for men and women warranted consideration. (Separate factor analyses for different age and educational groups may also be necessary; we anticipate following through these possibilities in future analyses.)

Kendall's tau beta was computed for each pair of indices. This particular statistic was used for two reasons. It avoids the necessity of collapsing categories of coding, as would have to be done if we employed tetrachoric correlations or phi coefficients; it also avoids the necessity of making unwarranted assumptions about units of measurement, as would be required if Pearsonian product-moment correlations were used. The logic of factor analysis is applicable to tau beta's since the statistic has been found to operate similarly to Pearsonian product-moment correlations usually used in factor analyses; the sampling errors and distributions of the two statistics are similar.[3]

[3] The authors would like to express their appreciation to William L. Hays and to John A. Sonquist for their help with these statistical problems and procedures.

The centroid method of factor extraction and IBM 650 Principle Axis Varimax program for rotation were used to obtain the final factor structures for men and women.

RESULTS AND DISCUSSION

Tables 2 and 3 present the matrices of intercorrelations for men and women, respectively. The significance of tau beta can be estimated: for the men ($N = 255$), tau's needed for significance at the .05 and .01 levels of confidence are .082 and .108, respectively; for the women ($N = 542$), the comparable figures are tau's of .054 and .091, respectively (cf. Kendall, 1948). Since the size of the samples involved here is large, and the size of the correlations required for significant deviations from zero is therefore relatively small, our major interest in these data is in the communalities of responses shown in the factor analysis.

In Table 4, the rotated factor loadings for both men and women are presented. The order in which the factors were extracted for men was 1,2,4,5,3; for women it was 1,3,2,4,5. In attempting to interpret the factor structures for men and women, we limited our consideration to those items which had factor loadings of .30 or higher.

In the main, the factor structures for men and women are similar; four parallel factors emerge and one additional factor appears for men. The similarity in these factor structures

TABLE 2

INTERCORRELATIONS (tau beta) OF 22 INDICES OF SUBJECTIVE ADJUSTMENT
(N = 255 employed married men with children[a])

	Worrying	Unhappiness	Future unhappiness	Nervous breakdown	Lack of uniqueness of self	Lack of self-acceptance	Shortcomings	Lack of strong points	Marital inadequacy	Marital unhappiness	Marriage problems	Negative orientation to children	Problems with children	Inadequacy as parent	Psychological anxiety	Physical ill health	Immobilization	Physical anxiety	Summary symptom score	Job dissatisfaction	Job problems	Job inadequacy
	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22
01		+19[b]	−03	13	−04	13	18	−01	10	13	05	01	03	07	28	01	06	09	16	−01	08	−04
02			04	10	07	−01	10	04	15	38	14	04	20	08	16	14	12	12	24	13	03	−09
03				06	07	09	−01	08	−08	10	−10	−06	01	−02	07	17	−02	08	07	−08	−03	−16
04					−07	07	09	01	10	03	14	−11	−01	27	29	14	16	29	33	03	23	08
05						−12	−07	11	02	05	−05	−04	−07	−08	−04	−02	−11	05	00	00	02	04
06							01	−08	06	−09	−06	02	03	00	28	02	12	06	13	05	−03	−01
07								06	15	14	07	00	10	18	06	−02	07	05	09	14	13	06
08									11	18	13	−04	03	09	02	03	01	06	03	08	−06	16
09										04	19	04	02	30	20	−01	24	10	20	03	26	07
10											23	−01	06	00	13	12	04	11	15	11	01	00
11												02	09	15	21	−01	19	05	17	08	13	06
12													06	−01	01	−17	09	−07	00	−01	00	−02
13														17	13	02	−01	12	09	10	10	05
14															15	06	14	17	23	10	25	18
15																16	34	32	57	12	12	05
16																	11	32	36	06	11	−11
17																		27	41	14	16	03
18																			58	06	15	−04
19																				14	22	−04
20																					04	14
21																						−06
22																						

[a] Includes only men responding to the one form of the interview on which all 22 items were asked. Ns for for specific correlations vary slightly due to incomplete data for some subjects.
[b] Plus sign is omitted from all other figures in table. All correlations are positive unless otherwise noted.

for men and women is an important finding; unexplained sex differences in obtained relationships are often the bane of the psychological researcher. However, the differences between men and women in the factor loadings of the various items on each factor are provocative. We highlight these in the discussion to follow.

Factor Structure for Men

Factor 1: Felt psychological disturbance. The variables that load highly on this factor are derived from the symptom checklist and questions concerning psychological and psychosomatic complaints (*psychological anxiety, immobilization, physical anxiety,* and *summary symptom score*). This factor represents an admission to symptoms that have particular implications for potential psychological dysfunctioning.

Since the variables that load highly on this factor all derive from a symptom checklist, filled out by the respondent himself, the question may arise as to what extent this factor represents a response set. It may be noted, in this connection, that the physical ill health symptom index, which also derives from this symptom checklist, is *not* highly loaded on this factor, which argues against a general "response set" interpretation.

It is noteworthy that this factor, entirely composed of symptom factor indices, does not have high loadings from other measures and further, that the symptom factor indices do not have high loadings on other factors, except on Factor 5. This, as we shall see, raises certain questions about the diagnostic use that has been made of symptom checklists.

Factor 2: Unhappiness. The general evaluations of *unhappiness* and *marital unhappiness* are the most highly loaded indices on this factor. In addition, the *problems with children* and *shortcomings* indices are also loaded on this factor. Thus "unhappiness" among men appears to be a general factor, tied to distress in several central life-areas—marriage, children, and the self-concept. While the loading of these different indices on

TABLE 3

INTERCORRELATIONS (tau beta) OF 19 INDICES OF SUBJECTIVE ADJUSTMENT
(N = 542 married women with children[a])

	Worrying	Unhappiness	Future unhappiness	Nervous breakdown	Lack of uniqueness of self	Lack of self-acceptance	Shortcomings	Lack of strong points	Marital inadequacy	Marital unhappiness	Marriage problems	Negative orientation to children	Problems with children	Inadequacy as parent	Psychological anxiety	Physical ill health	Immobilization	Physical anxiety	Summary symptom score
	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
01		+18[b]	−05	16	−04	10	05	−01	08	05	02	01	06	04	19	06	06	08	14
02			−04	11	02	02	00	04	−01	44	24	06	05	03	19	10	02	15	18
03				−01	05	−03	00	08	−12	01	−02	−03	−09	−02	−06	01	−10	05	01
04					−07	13	08	01	14	16	18	−01	20	12	32	21	20	31	45
05						−04	−05	13	−01	02	−04	−06	−05	04	−06	00	−01	−04	−03
06							07	07	14	02	01	−00	08	12	11	03	07	10	08
07								−03	20	04	10	03	07	16	08	−03	14	07	08
08									01	04	00	−03	−06	−01	01	07	−04	03	−01
09										01	14	03	15	37	15	−00	16	14	16
10											28	07	07	04	15	06	01	18	20
11												08	13	18	15	−06	08	14	19
12													04	−01	01	−04	07	08	07
13														17	14	−01	03	06	10
14															16	00	15	13	16
15																24	23	32	58
16																	07	25	38
17																		22	33
18																			58
19																			

[a] Includes only women responding to the two forms of the interviews on which all 19 items were asked. Ns for specific correlations vary slightly due to incomplete data for some subjects.
[b] Plus sign is omitted from all other figures in table. All correlations are positive unless otherwise noted.

TABLE 4

FACTOR LOADINGS FROM NORMALIZED VARIMAX ROTATIONS[a]

Indices of subjective adjustment	Factor 1		Factor 2		Factor 3		Factor 4		Factor 5		h^2[b]	
	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women
Worrying	+18[c]	13	28	−10	−08	−02	−26	−07	07	−27	19	11
Unhappiness	14	11	56	−66	04	07	09	01	−08	−14	35	47
Future unhappiness	00	02	08	00	13	08	−11	28	−32	07	14	09
Nervous breakdown	29	41	02	−13	−39	−16	−12	−06	−19	−35	29	34
Lack of uniqueness of self	−02	−03	04	−02	10	−02	31	33	−04	09	11	12
Lack of self-acceptance	26	08	03	02	07	−18	−40	03	00	−22	23	09
Shortcomings	00	05	31	−03	−28	−34	−07	−06	08	−02	19	12
Lack of strong points	06	02	18	−02	−05	−01	31	34	−02	−06	13	12
Marital inadequacy	28	10	12	01	−37	−58	09	−05	21	−11	28	36
Marital unhappiness	10	11	54	−64	07	−01	19	06	−09	−05	35	43
Marriage problems	21	07	20	−42	−22	−26	21	−06	14	00	20	25
Negative orientation to children	04	05	06	−14	12	−05	−05	−19	23	15	08	08
Problems with children	−03	02	31	−08	−16	−22	−06	−18	−05	−27	13	16
Inadequacy as parent	13	08	11	−06	−59	−56	05	05	−02	−09	38	33
Psychological anxiety	70	57	24	−14	−09	−11	−19	−10	−05	−28	59	44
Physical ill health	18	46	08	02	−05	14	06	12	−55	−15	35	27
Immobilization	53	36	04	01	−17	−26	02	−14	02	06	31	23
Physical anxiety	42	62	07	−19	−17	−15	07	03	−49	01	45	44
Summary symptom score	66	85	14	−14	−19	−12	00	00	−37	−06	64	76
Job dissatisfaction	15		20		−09		14		02		09	
Work problems	11		01		−44		01		−10		22	
Job inadequacy	08		−02		−15		28		20		15	

[a] Based on samples of 255 men and 542 women.
[b] h^2 is equal to the communality of the item.
[c] Plus sign is omitted from all other figures in table. All factor loadings are positive unless otherwise noted.

an unhappiness factor is not surprising, it is interesting to note that a quite different pattern appears for women where, as we will see, only the marital indices are loaded on the unhappiness factor.

Factor 3: Social inadequacy. This factor seems to represent the self-questioning that goes on when a man considers himself in relationship to his social world. Some men entertain considerable self-doubt in their role relationships with others; others are self-confident or refuse to admit to doubt. Items loaded on this factor seem to be measuring the extent to which a man asks himself: Am I doing a good job at the things I am doing—worker, husband, father? Such feelings are objectified into *job problems* (.44), *inadequacy as a parent* (.59), and *marital inadequacy* (.37).

If this factor is viewed as measuring general social inadequacy, the relative weights of the factor loadings of these indices suggest that feelings of inadequacy as a father may be the best indications of a man's general feelings of inadequacy, or at least may be better measures of this factor than feelings of inadequacy as a husband or as a worker. The difference between the loadings of the parental and marital indices is particularly interesting. It suggests that admission to inadequacy may involve minor rather than major deviance from ideal role behavior more often when one is thinking of one's behavior as a husband than when one is thinking of one's behavior as a father. In marriage—perhaps because of the greater complexity of the role—we find built-in allowances permitting deviance for men. This is rather graphically illustrated in the almost jocular sanction given to minor inattentiveness on the part of the husband to his wife—forgetting anniversaries, not helping with the dishes, going out with the boys. Being a father does not seem to provide such culturally approved safety valves, perhaps because the demands of the role are less complex—supporting the child being such an all important feature of the role. Consequently admission to inadequacy as a father may be diagnostic of more serious feelings of social inadequacy than admission to inadequacy as a husband.

Also loaded on this factor is the experience, some time in one's life of having had or been

close to a "nervous breakdown." For a man, a nervous breakdown may be a crucial indication of collapse in the face of pressures in the outside world, an extreme disruption in his ability to face his life obligations.

In this factor, as in the others, the investigation of mental health through the layman's self-perceptions has yielded a cluster of indices that has a parallel in the "experts'" attempts to establish criteria for mental health. One of the six major criteria for mental health that Jahoda (1958) describes is that of "environmental mastery." Included in this category is the person's adequacy in a variety of life situations—love, work, play, interpersonal relations. While this criterion is usually evaluated from the "objective" observer's point of view, the present results indicate it is an important aspect of self-evaluations as well.

Factor 4: Lack of identity (versus heightened self-awareness). In the two factor analyses this is the only factor which emerges in a bipolar fashion (that is, with high loadings in both the negative and positive direction). On the one pole are found *lack of uniqueness of self*, together with *lack of strong points in the self;* on the other pole we find *lack of self-acceptance.* This at first seems like a perplexing factor, but results stemming from our previous analysis of self-perception in the national survey, suggest a possible explanation of the factor (Gurin et al., 1960, Ch. 3). As we noted in that study, responses to questioning about the self are influenced both by how introspective a person is and by how positively or negatively he evaluates himself. Furthermore, there was some evidence that these two determinants are interrelated, that tendencies toward introspection and self-awareness are related to feeling negative about the self. Being able to see negative traits in the self implies a certain degree of introspection.

In the light of these previous findings, we interpret one pole of this factor to be this heightened self-awareness, negatively loaded primarily with the lack of self-acceptance index. Lack of self-acceptance was coded when the respondent reported any negative aspects of the self, something which, as we have indicated, we interpreted as a particular indication of self-awareness. On the other pole we have an opposite reaction—a rejection of looking into the self, a rejection of considering the self as an object of appraisal, what we might call a lack of identity. The indices positively loaded on this pole of the factor are lack of strong points in the self and lack of uniqueness of self (for example, "There is nothing good about me," "I'm not different from anyone else"). Lack of strong points, in short, at least in its loading on this factor, is not so much a "negative" perception of the self as it is an indication of an inability to introspect.

This factor is of particular interest because theoretical discussions of positive mental health have tended to link self-awareness and positive feelings about the self: "a healthy person knows who he is and does not feel basic doubts about his inner identity" (Jahoda, 1958, p. 29). Without denying this, it is also true (as Jahoda has also pointed out) that these "healthy" self-attitudes at the same time include the ability to see and realistically accept the negative aspects of the self. And it is difficult at times to determine when the "healthy" relationship between self-awareness and realistic self-appraisal turns into a relationship between self-awareness and a negative view of the self. This was illustrated in the previous analyses of the data in this study, where, in comparing the responses to the self-percept questions among the different demographic subgroups of the population, it was noted that ". . . the social and cultural conditions that lead to introspection and a heightened self-awareness may also lead not only to realistic self-criticism, but to a generally negative self-percept" (Gurin et al., 1960, p. 83).

Factor 5: Physical distress. Some of the indices based on the symptom checklist again stand out in this factor—*physical ill health*, physical anxiety, and summary symptom score. This factor thus seems to be defined by physical dysfunctioning. An important sidelight here is that the highest loading that the *future unhappiness* index has is on this factor (.32). A man's optimism or pessimism about his future is tied in large measure to his potential as a worker and economic provider. Physical disability in a man is a major con-

dition interfering with this capacity. Thus it is reasonable that anticipation of future happiness or unhappiness might draw its heaviest loading on this factor for men.

A comparison between Factors 1 and 5 is of interest because of the considerable overlap in the indices loaded on them. These two factors account for the major portion of the variance of the indices based on the symptom checklist. Physical anxiety and summary symptom score are highly loaded on both factors. In contrast, the two most psychological symptom indices—psychological anxiety and immobilization—appear only on Factor 1, and the most physical symptom index—physical ill health— appears only on Factor 5. This points to a general as well as two specific components of the symptom checklist for men. Some implications of this distinction may be clarified when we examine the factor structure for women, where we will see quite a different structure for these symptoms.

Factor Structure for Women

Five factors were also extracted for women. Except for one of them, there seem to be some direct parallels between the two factor structures. Although there are differences which suggest ways of conceptualizing male and female differences in psychological functioning, the most striking over-all impression is that the two sets of data are very much alike. The four similar factors in the factor structure of both men and women are: felt psychological disturbance, unhappiness, social inadequacy, and lack of identity. There is no factor in the structure for women comparable to the physical distress factor found in the data from men's responses, and the fifth factor for women is difficult to interpret; only one index loads highly on it—*nervous breakdown*.

In what ways do the factors for females seem to differ from males? Let us consider each in order. There are many different theoretically important dimensions underlying sex as an analytic variable and each might have implications in the present analysis. However, we will select interpretations of differences that seem most plausible but recognize that other interpretations are possible.

Factor 1: Felt psychological disturbance. This factor is both similar and dissimilar to Factor 1 for men. Four indices are highly loaded on this factor for both men and women: psychological anxiety, immobilization, physical anxiety, and summary symptom score. But the physical ill health index is highly loaded on this factor for females, whereas for males it was part of a physical distress factor (Factor 5), which we do not find in the factor structure for the females. Furthermore, for men physical anxiety and summary symptom score are not uniquely loaded on Factor 1, as they are for women. Thus it seems that women do not separate admission to bodily complaints from admission to complaints about their general psychological condition; men evidently make this distinction. Perhaps these results are tied to the greater tendencies among women to introspect and to see their problems in psychological terms, in contrast to men's greater tendencies to externalize their problems.

Factor 2: Unhappiness. As was observed in the discussion of this factor for men, general unhappiness and marital unhappiness seem to go together, both being highly loaded on this "unhappiness" factor. However, certain differences between men and women also appear. Among women this factor does not have high loadings on two indices—shortcomings and problems with children—that were highly loaded on this factor for men. And among women this factor does have a high loading on one index—*marital problems*—which did not appear among the indices with high loadings on this factor for men.

Among women, then, the unhappiness factor, with high loadings on general unhappiness, marital unhappiness, and marriage problems, seems more exclusively tied to marriage than it was for men. This may be taken as some support for the hypothesis that marriage is more central to a woman's life, and hence plays a more significant role in her general happiness.

But why are not the shortcomings and problems with children indices also included in this factor? With respect to the shortcomings index particularly, one might have expected that marriage, being more central in a woman's life, would be more rather than

less tied to self-evaluation and self-criticism, and that for women a factor heavily loaded on marital unhappiness would show higher rather than lower loadings on a shortcomings index. But one of the findings reported in the larger study may suggest why this is not so. In the analysis of adjustment to the marital role, it was noted that women blame their husbands for marital difficulties more often than men blame their wives (Gurin et al., 1960). This difference in attribution of blame seemed to relate to the male-female role distinctions in marriage. The fact that the woman's role in marriage is more passive and receptive than the man's apparently more often enables her to separate marital failure from failure in the self.

The activity-passivity dimension in role behavior may also help explain why marital unhappiness in women also seems to be separate from distress over the children. Only in this case it is the active role that a woman plays as a mother and the contrast with the more receptive and passive role as a wife that might explain why problems in both areas are not loaded on the same factor, why a woman can apparently separate her reactions to marriage from her response to her children. Unhappiness in the more passive role as wife may actually lead to an overly positive involvement in the more active role as mother—a phenomenon which has often been observed in women who, unhappy and feeling helpless in a frustrating unfulfilling marriage, attempt to compensate by an overinvestment in their children.

Factor 3: Social inadequacy. The hypothesis that the marriage is more important for the feelings of adjustment for women than for men is given further support in a sex difference obtained in the social inadequacy factor. The loading on this factor for marital inadequacy is .58 for women and .37 for men. Both men and women have very high loadings on the inadequacy as parent index. One further kind of self-questioning, the shortcomings index, is also highly loaded (.31) on this factor for women; this is again similar to the finding for men, where this index falls just short (.28) of our criterion for consideration.

Being a mother and being a wife are both central to a married woman's identity; feeling inadequate in either role is equally important in a woman's feelings of inadequacy and failure. Neither role apparently provides the culturally allowed safety valves for deviance that were noted earlier in relation to the husband's role.

Factor 4: Lack of identity. Only one pole of the comparable factor for men emerged for women: Lack of uniqueness of self and lack of strong points in the self again emerged together in this factor, but lack of self-acceptance does not load on the other pole of the factor for women. What might be suggested is that women generally tend to be more introspective, so that a heightened self-awareness is not a special factor for them.

Implications of the Factor Analysis

The multiple criterion approach to the delineation of a person's adjustment that has received increasing support in recent writings is supported by these factor analyses. Even in that limited evaluation of mental health that is available from what a person tells us as he describes his own adjustment, whether a person gives an "adjusted" response or a "maladjusted" response depends on what the question is. Evidently people do not respond to a series of questions covering life adjustment with one general evaluative framework which one could call the person's subjective adjustment. Rather, a person seems to respond to these series of questions from a number of different frameworks, perhaps represented by the factors that emerged from the factor analysis. There are undoubtedly some people who approach all of the questions about adjustment in a general "distressed" way, for whom being high in one of these factors implies being high in all. In his work with people who seek psychological help the clinician perhaps most often encounters this general discontent that transcends all dimensions of self-description—but this pattern of self-evaluation may be less common among the general population.

It is particularly interesting to observe that the distinctiveness of the factors applies to the symptom checklist as well as to the other adjustment measures that were studied. Since symptom checklists such as the one on which the present symptom indices are based, are often used as a quick psychiatric screening

technique, one might have expected the symptom indices to represent a general factor, highly loaded on all the factors. But this is not so. For both men and women one factor (Factor 1) is composed entirely of symptom indices. And although these symptom items do have some loadings on the other factors, these loadings are not high for women, and for men are high only on Factor 5, which is itself largely composed of the physically loaded symptom indices. Apparently, the major aspects of distress included in a symptom checklist, although partial components of other kinds of psychological distress experiences, are fairly distinct from them. Implicitly, then, the judgments that are based on utilizing such symptom batteries as a diagnostic tool seem to be minimizing the importance of other aspects of feelings of maladjustment. Essentially, they single out one from a number of possible mental health criteria.

This reliance upon one single criterion—whether of symptoms or other factors—has special significance when attempts are made to compare different subgroups of the population in mental illness or health. Such a comparison is questionable if these subgroups vary in their tendencies to experience distress as symptoms or in other ways. If subgroups do vary in the ways in which they experience distress, then the "mental health" evaluation of the different groups would vary according to the particular component of mental health that was chosen as the basis for evaluation; for example, a "symptom" criterion would provide a different ranking of groups than would be obtained from a "felt inadequacy" criterion.

That this does occur has been demonstrated in prior results reported by Gurin et al. (1960). Those results were based on the separate questions used in the present factor analysis. The general point can be illustrated better now that the results of the factor analyses are available. A systematic investigation of subgroup differences in responses to the indices based on the several factors is contemplated. For the present, we can illustrate this problem by considering the responses of groups differing in education.

In order to facilitate these comparisons, factor scores were computed for each of the factors for men and women. The factor scores were based on summations of those indices that were both highly loaded (.30 or more) on the factor in question *and* were not highly loaded on any other factor. These criteria were established in order to yield factor scores that would both be clearly representative of the respective factors and would be as independent of one another as possible.[4] In the summations, each index was given equal weighting.

The indices comprising each factor score are listed below:

Factor 1. For men: psychological anxiety and immobilization; for women: summary symptom score.[5]

Factor 2. For men: unhappiness, shortcomings, marital unhappiness, and problems with children; for women: unhappiness, marital unhappiness, and marriage problems.

Factor 3. For men: nervous breakdown, marital inadequacy, inadequacy as parent, and job problems; for women: shortcomings, marital inadequacy, and inadequacy as parent.

Factor 4. For men: lack of uniqueness, lack of self-acceptance,[6] and lack of strong points; for women: lack of uniqueness and lack of strong points.

Factor 5. For men: future unhappiness and physical ill health.

The relationships between education and these factor scores are presented in Table 5. The expectation, based on previous results, that the relationship between education and the experience of distress would vary for the

[4] Since no indices were loaded exclusively on only one factor, complete independence of factor scores was not obtained. But among the 16 possible intercorrelations (10 among the five factors for men and 6 among the four factors for women) 12 yielded tau's of .10 or less. The four exceptions were tau's of .13 between Factors 1 and 2 for men, .24 between Factors 1 and 3 for men, .25 between Factors 1 and 2 for women, and .20 between Factors 1 and 3 for women.

[5] The summary symptom score was used as the sole basis for computing the Factor 1 score for women both because of its extremely high loading on the factor (.85) and because most of the items comprising the other symptom indices loaded on this factor are also included in the summary symptom score.

[6] Because of the bipolar nature of this factor, the rankings for the lack of self-acceptance index presented in Table 1 were reversed in computing the factor score.

TABLE 5

EDUCATION AND FACTOR SCORES BASED ON SELF-EVALUATIONS OF DISTRESS (within-sex)

Factor	Score[a]	Men			Women		
		Grade school[b]	High school	College	Grade school	High school	College
1. Felt psychological disturbance	H	18%	16%	23%	35%	21%	13%
	MH	21	19	25	25	23	30
	ML	22	24	19	18	28	23
	L	39	41	33	22	28	34
	N[c]	85	112	57	139	315	82
2. Unhappiness	H	35%	29%	28%	32%	17%	17%
	MH	33	24	20	32	27	22
	ML	13	19	28	22	29	33
	L	19	28	24	14	27	28
	N	70	97	50	139	317	82
3. Social inadequacy	H	19%	22%	38%	13%	21%	24%
	MH	19	25	30	19	26	35
	ML	35	24	18	24	25	26
	L	26	29	14	44	28	15
	N	68	96	50	109	263	74
4. Lack of identity	H	15%	6%	2%	7%	5%	7%
	MH	41	34	32	{36	{31	{34
	ML	27	45	40			
	L	17	15	26	57	64	59
	N	84	111	57	138	312	83
5. Physical distress	H	38%	11%	10%			
	M	35	31	39			
	L	27	58	51			
	N	63	101	49			

[a] The scores represent the closest approximation into quartiles that was possible except for Factor 4 for women and Factor 5 for men where only thirds were possible. The abbreviations H, M, L stand for high, medium, and low, respectively.
[b] Categorization based on highest educational level begun.
[c] *N*s vary due to incomplete data on items comprising the several factor scores.

several factor scores is clearly confirmed by the data presented in this table. The less educated men would be judged as indicating *less* distress than the more educated ones if the felt psychological disturbance (Factor 1) and particularly the social inadequacy (Factor 3) scores were considered. But, they would be judged as *more* distressed if the unhappiness (Factor 2), lack of identity (Factor 3), or physical distress (Factor 5) scores were considered. It is interesting to note that some of these findings—specifically those relating to psychological disturbance, social inadequacy, and physical distress—support much of the thinking that has been done with respect to social class differences in the expression of psychological distress. Using education as an index of social status, the findings conform to hypotheses that have been offered about psychological symptoms and intrapunitiveness being more common to the middle class, and bodily symptoms being a more characteristic working class mode of expression.

Women with differing levels of education also respond differentially to the several categories of self-evaluations of distress represented by the factor scores. The less educated women indicate *more* felt psychological disturbance (Factor 1) and unhappiness (Factor 2), but *less* social inadequacy (Factor 3) than the more educated women. Education is unrelated to Factor 4—lack of identity—for women.

As can be seen, these relationships differ in some ways from those for men. Particu-

larly, they point up a difference between the symptom factors for men and women. The computation of factor scores in a manner that accentuated the uniqueness of each factor had the effect of emphasizing the different determinants of the responses to symptom checklists for men and women. For men, there appear to be somewhat distinct psychological and physical components involved in the responses, as reflected both in the emergence of Factor 5 and in the differential relationships of Factors 1 and 5 with education; for women, only one symptom factor emerged. The use of a symptom checklist with women, like the summary symptom score which is the basis for their Factor 1 scores, clearly leads to the conclusion that the less educated groups are more distressed. For men this negative relationship is obtained only for the Factor 5 scores, where only the physical components of the symptom checklist (physical ill health) are included in the factor score. When the factor score is based on the more psychological components of a symptom checklist (Factor 1), a clear relationship with education is not obtained; the relationship is slightly positive, with college educated men scoring somewhat higher than grade school educated men. The multiple criterion point, however, remains true for both men and women: the selection of a symptom checklist as the more significant indicator of mental health means that certain groups will be selected as "sick" or "healthy"; a different selection would be made if other criteria for mental health were stressed.

We have highlighted the fact that *different distinctive factors* emerge in this factor analysis of self-evaluations of distress, and have used the analysis of educational subgroup differences to support the need for a multiple criterion approach to subjective adjustment. It is still possible, however, that one might find one factor or perhaps a pattern of the separate factors of self-evaluation which, through a method of external validation, might be chosen as *the* best indicator of mental illness or health. But the problem in such an attempt lies in determining what the external criterion should be. It is difficult to choose such an external validation criterion given the considerable conceptual disagreement as to what the definitions and criteria of mental illness and health should be. Some of these difficulties are illustrated in the attempt to single out symptom lists as the diagnostic tool, with psychiatrists' judgments being the external validating criterion. As we have seen, this seems to minimize certain aspects of psychic dysfunctioning which many would agree should be included in any psychiatric judgment of healthy or neurotic functioning.

Although the problem is partly a function of the imperfections of our knowledge in this area, it is likely that the many different relevant criteria of disturbance can never be encompassed within a single all-embracing criterion. In attempting to predict to some external criterion from the kinds of factors differentiated in this study, it is probably more appropriate to think of different factors being relevant for different predictions than to think in terms of a single criterion and prediction. For example, if we wish to predict the type of disturbance that seems to be implicit in psychiatrists' diagnoses, subjective evaluations that involve a heavy component of psychophysical symptomatology would be the best predictors. On the other hand, for predicting the kind of impairment that leads to immobilization, breakdown, and the self-diagnosis involved in going for help with one's problems we would look to a quite different set of subjective adjustment measures —for example, those heavily loaded on the inadequacy factor. And for a prediction of "positive" mental health still other factors might be most relevant.

In comparing the factor structures of the men and women in our study, both the similarities in the structures and the differences can have important implications. The communality between the two factor structures points to the important psychological dimensions of subjective distress that can transcend social roles and cultural subgroups. If one were looking for the basic underlying psychological parameters of distress, this focus on the communality would be of greatest importance. A logical step in such a focus would involve factor analyses on specific social groups—different age groupings, different educational statuses, different income levels—to pinpoint even further those communalities that exist in all groups.

What are the implications of finding dif-

ferences in factors for men and women? Although both sexes may encounter similar problems in growing up and in current living, they undoubtedly have learned to deal with these problems in different ways. Differences in factors for men and women may reflect critical ways in which the sexes differ in handling or expressing psychic difficulties. For example, since the physical ill health symptom index appears as part of the felt psychological disturbance factor for women but is separate for men, one perhaps has support for the view that the externalization of psychic disturbance into somatic distress is more common for men than for women. Analyses which focus on such differences in factors can pave the way to a better delineation of the critical differential experiences of men and women, or of any other groups so compared.

SUMMARY

A representative sample of the resident population of adults in the United States was interviewed about their feelings of distress in different life areas: perceptions of the self; symptoms of distress; adjustment in marriage, parenthood, and work. Nineteen indices were constructed for special subpopulations —married women with children and employed married men with children. In addition, three indices related to aspects of work adjustment were used only for these men. The indices were intercorrelated and the intercorrelations were factor analyzed separately for men and women.

Five distinctive factors emerged for men and four for women. There was considerable apparent overlap between the two factor structures. For both men and women the factors were identified as: felt psychological disturbance (Factor 1); unhappiness (Factor 2); social inadequacy (Factor 3); lack of identity (Factor 4). For men the fifth factor was labeled physical distress.

The emergence of several distinct factors underlying self-evaluations of distress was discussed as reflecting the need for a multiple criterion approach to the definition of mental health. In particular it was noted that findings about differential distress in population subgroups depend upon the particular syndrome of distress that is used as the criterion of mental health; this point was illustrated for educational subgroupings.

The overlapping factor structures of men and women point to basic dimensions of distress applicable to all people. The differences in the factor structures for men and women point to major sex differences in the experience of distress. In particular, physical distress seems to be a distinctive factor for men while it is tied to psychological symptoms for women.

REFERENCES

EATON, J. W. The assessment of mental health. *Amer. J. Psychiat.*, 1951, **108**, 81–90.

FIEDLER, F. E., DODGE, JOAN S., JONES, R. E., & HUTCHINS, E. B. Interrelations among measures of personality adjustment in nonclinical populations. *J. abnorm. soc. Psychol.*, 1958, **56**, 345–351.

GURIN, G., VEROFF, J., & FELD, SHEILA. *Americans view their mental health*. New York: Basic Books, 1960.

HERZBERG, F., MAUSNER, B., & SNYDERMAN, BARBARA. *The motivation to work*. New York: Wiley, 1959.

INKELES, A. Industrial man: The relation of status to experience, perception, and value. *Amer. J. Sociol.*, 1960, **66**, 1–31.

JAHODA, MARIE. Toward a social phychology of mental health. In J. M. E. Senn (Ed.), *Symposium on the healthy personality*. New York: Josiah Macy, Jr. Foundation, 1950. Pp. 211–230.

JAHODA, MARIE. The meaning of psychological health. *Soc. Casewk.*, 1953, **34**, 349–354.

JAHODA, MARIE. *Current concepts of positive mental health*. New York: Basic Books, 1958.

KENDALL, M. G. *Rank correlation methods*. London: Griffin, 1948.

KISH, L. Selection of the sample. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences*. New York: Dryden, 1953. Pp. 175–240.

MACMILLAN, A. M. The health opinion survey: Technique for estimating prevalence of psychoneurotic and related types of disorder in communities. *Psychol. Rep.*, 1957, **3**, 325–329.

MORSE, NANCY C. *Satisfactions in the white-collar job*. Ann Arbor: Survey Research Center, 1953.

ROGERS, C. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.

SCOTT, W. A. Research definitions of mental health and mental illness. *Psychol. Bull.*, 1958, **55**, 67–87.

SMITH, M. B. Optima of mental health: A general frame of reference. *Psychiatry*, 1950, **13**, 503–510.

SMITH, M. B. Research strategies toward a conception of positive mental health. *Amer. Psychologist*, 1959, **14**, 673–681.

TERMAN, L. M., & WALLIN, P. The validity of marriage prediction and marital adjustment tests. *Amer. sociol. Rev.*, 1949, **14**, 497–505.

(Received February 8, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 64, No. 3, 206-214

DIFFERENCES AND DISAGREEMENT AS FACTORS IN CREATIVE GROUP PROBLEM SOLVING[1]

L. RICHARD HOFFMAN, ERNEST HARBURG, AND NORMAN R. F. MAIER

University of Michigan

Differences in opinions and ideas among group members need not lead to dissatisfactions and unpleasant experiences but rather can lead to invention and creative problem solving (Maier, 1958). When opposing viewpoints are brought together in a free exchange, the resulting conflict may be resolved by the creation of a new solution which either reorganizes or incorporates the different viewpoints. This solution becomes a new meeting ground and is therefore satisfactory to all. Recent studies have provided support for these propositions by showing that differences among group members promote effective problem solving (e.g., Ghiselli & Lodahl, 1958; Hoffman, 1959; Hoffman & Maier, 1961a).

The mere existence of different viewpoints in a group does not guarantee their full expression or consideration, however, especially if one of the viewpoints is held by a person of authority over the other group members. An authority figure is likely to dominate the discussion and force his ideas on the group, disregarding or even suppressing the opposing ideas of his subordinates (Maier & Hoffman, 1960, 1961; Torrance, 1955). His dominance interferes with the free expression of differing opinions and reduces the possibility of the emergence of creative solutions to problems.

The usual approach to overcome the dominance of a person in authority over discussions is to attempt to reduce his influence in some way. Maier (1952) has suggested that group leaders refrain from suggesting ideas, provide expert information as needed, and allow subordinates to make all suggestions for problem solutions. In this way the leader's favored idea is less likely to dominate the discussion than if he makes his views known before his subordinates have had their chance to be heard.

A less common way of ensuring the expression of differing opinions by subordinate members of the group is to increase their commitment to those opinions. Subordinates are likely to give strong expression to their viewpoints when they feel sufficiently secure to disagree with the person in authority, as when they belong to a strong labor union, or when they have sufficiently strong feelings about the subject being discussed. Increasing the subordinates' commitment so that their point of view can effectively oppose that of the authority figure, thus may be regarded as a way to introduce conflict and to provide the conditions for creative problem solving.

These two approaches to ensuring the free expression of different opinions in the group —increasing the subordinates' commitment and reducing the superior's dominance—were combined in the present experiment to determine their separate and combined effectiveness in improving the quality of group solutions to a problem.

The Change of Work Procedure problem (Maier, 1952)—a role playing situation involving a conflict between a foreman and his subordinate workers about the relative merits of two work methods—was selected for use in this experiment, since it seemed to provide the conditions needed to test the experimental variations. Experience with this problem in its standard version has shown that the foreman's preferred work method is adopted in a majority of the cases and creative solutions occur relatively infrequently. Previous research also has indicated that the production of creative solutions to the problem was increased when subjects' orientation to authority relations was low (Maier & Hoffman, 1961), when foremen were trained to be permissive (Maier, 1953), and when the groups solved the problem a second time (Maier &

[1] This investigation was supported by a USPHS research grant (M-2704) from the National Institute of Mental Health, United States Public Health Service.

Hoffman, 1960). The improvement in the last study seemed to result principally from the foreman's greater willingness to permit the workers to influence the second decision, since a solution had been reached in the first meeting.

Variations in the relative dominance of the foreman and in the strength of the workers' feelings were introduced in the present study by altering the degree of motivational support for each side's position, as described in the role instructions given to each group member. The workers were provided with reasons to be either more or less committed to the work method they were currently using. The foremen were given the basis either for strongly encouraging the adoption of a new work method or for being more considerate of the workers' feelings in suggesting its adoption.

If conflict is potentially conducive to effective problem solving, high quality solutions should be produced more often by groups where the workers are more committed to the existing work method than where they are less committed, since conflict is more likely to occur in the former condition than in the latter. Conflict can also lead to personal animus instead of invention, especially if the foreman does not see conflict as offering the possibility of invention. High quality solutions are thus more likely to appear where the foreman is considerate in the face of strongly committed workers than where the foreman is himself strongly committed to his own suggestion.

While an initial conflict is hypothesized to set the stage for possible creativity, whether the final result will be a creative solution or acceptance of one of the less satisfactory solutions will still depend upon the group's ability to discover a creative solution. Even when the foreman is willing to consider alternative solutions, these cannot be adopted unless they are developed by someone. Differences among the group members appear to be important here also, in that they provide a variety of perceptions of the problem and, therefore, different "directions" in thinking (Maier, 1930). A variety of directions is postulated to increase the probability of discovering a creative solution to a problem.

Mixing the sex composition of a group is one way of providing such differences. In a previous study in which the Change of Work Procedure problem was used (Hoffman & Maier, 1961a), mixed-sex groups produced a somewhat (but not significantly) higher proportion of high quality solutions than did all-male groups, despite the fact that all-female groups are usually poor at problem solving. In the present experiment the sex composition of the groups was varied systematically so that the quality of solutions of mixed-sex, all-male, and all-female groups can be compared under each condition of foreman-worker conflict. Mixed-sex groups may be expected to produce more high quality solutions than single-sex groups. Since the availability of different perspectives or directions should be especially valuable where conflict leads to a search for new solutions, the mixed-sex groups may be expected to be especially effective where the workers in the groups are committed to the existing work method, i.e., where conflict occurs.

METHOD

Problem

The Change of Work Procedure problem (Maier, 1952) is a role playing situation involving a foreman and three of his subordinate workers in an automobile factory. The workers assemble fuel pumps and are paid on a group piece-rate basis. The assembly operation is divided into three simple jobs and the men presently rotate hourly among the three positions. The role playing consists of a meeting called by the foreman to discuss a change in work method in which each worker would work on only one position, his best according to time-study data provided for the foreman in his role instructions.

In a majority of groups, the foreman is able to overcome the workers' fears and resistance and to sell the group on the virtues of the *new* fixed-position method—principally that the increased production achieved would result in an increase in their pay. In some instances the workers refuse to change and retain the *old* method of working, usually on the grounds that the boredom of working on the new method would cause a decline, or at least no increase, in production. In a still smaller percentage of groups, a third type of work method is produced —a type in which the advantages of the individual differences in worker productivity are exploited, yet boredom is still relieved through rotation. These solutions have been termed *integrative*. Their production represents the release of the group from a situation of choice between the old and new methods to permit a search for additional solutions to

the problem. It is the integrative solutions that are here considered to be evidence of creative problem solving by the group.

Establishment of Experimental Conditions

Conflict. The degree of conflict was varied by altering parts of the standard role instructions to the foreman and workers.

All foremen received the following instructions:

> You are the foreman in a shop and supervise the work of about 20 men. Most of the jobs are piece-rate jobs and some of the men work in teams and are paid on a team piece-rate basis. In one of the teams, Jack, Walt, and Steve work together. Each one of them does one of the operations for an hour and then they exchange, so that all men perform each of the operations at different times. The men themselves decided to operate that way and you have never given the plan any thought.
>
> Lately, Jim Clark, the methods man, has been around and studied conditions in your shop. He timed Jack, Walt, and Steve on each of the operations and came up with the following facts:

TIME PER OPERATION

	Position 1	Position 2	Position 3	Total
Jack	3 minutes	4 minutes	4½ minutes	11½ minutes
Walt	3½ minutes	3½ minutes	3 minutes	10 minutes
Steve	5 minutes	3½ minutes	4½ minutes	13 minutes
				34½ minutes

> He observed that with men rotating, the average time for all three operations would be ⅓ of the total time or 11½ minutes per complete unit. If, however, Jack worked in the Position 1 spot, Steve in the Position 2 spot, and Walt in the Position 3 spot, the time would be 9½ minutes, a reduction of over 17%. Such a reduction in time would amount to a saving of more than 80 minutes. In other words, the lost production would be about the same as that which would occur if the men loafed for 80 minutes in an 8-hour day. If the time were used for productive effort, production would be increased more than 20%.

In addition, to encourage commitment to the new work method, half the foremen (referred to hereafter as Dominant foremen) were told in their instructions:

> You were very enthusiastic about this idea since it seemed to offer the men an opportunity to make more money without any more effort. At the same time, it will yield higher production for your unit. Although you have been satisfied with their production up to now, you have called Jack, Steve, and Walt in to tell them about this new plan. You see no reason why they should object to this change in operation.

The rest of the foremen (referred to hereafter as Considerate foremen) were told instead:

> This argument made pretty good sense to you so you have decided to take it up with the men. Since their production has been satisfactory up to now, you don't want to make any changes without their approval.

Although these latter instructions were designed to make the foreman favorable to the new method, they also should have made him more considerate of the workers' feelings on the matter.

Parts of the workers' roles were altered also to make them feel more or less bored with the present situation as the conditions required, thus more or less committed to the present work method (the old solution). The workers whose roles made them more committed to the old work method will be called Strong workers and those who were intended to be less committed will be called Weak workers to represent the strength of the feelings involved.

Some examples of the specific changes made in the role instructions will give the reader a sense of their nature. An example of the Weak and Strong instructions is given for each character.

> (Weak) Sometimes the work gets to be monotonous.
>
> (Strong) The work is very monotonous.
>
> (Weak) The three of you exchange positions every hour.
>
> (Strong) Because the work is so monotonous, the three of you decided to exchange positions every hour.
>
> (Weak) You have learned to set your pace to music you hum to yourself and the jobs don't bother you at all.
>
> (Strong) You have tried to make them [the jobs] less monotonous by talking, daydreaming, and changing your pace. You have even tried to set your work pace to music you hum to yourself. Each of these seems to help for a short time. . . . the change adds variety to the work.

In addition to these feelings that the workers hold about the monotony of the job, Steve is afraid of losing his job because he is inefficient and Walt is antagonistic toward the time-study man because he feels time study is a waste of the company's money.

Four experimental conditions varying in their degree of conflict were thus developed by combining the two forms of the foreman's role and the two forms of the workers' roles: Dominant foreman-Strong workers; Considerate foreman-Strong workers; Dominant foreman-Weak workers; Considerate foreman-Weak workers.

Sex Composition. The sex composition of the groups was varied systematically within each conflict condition by assigning eight groups each of all-male, all-female, and mixed-sex (two males and two females) randomly to each conflict condition. Thus

the 96 four-person groups were assigned to the following conditions with eight groups in each cell:

Foreman:	Dominant		Considerate	
Workers:	Strong	Weak	Strong	Weak
Sex Composition:	Male Mixed Female	Male Mixed Female	Male Mixed Female	Male Mixed Female

Also, in each conflict condition the foreman's role was assigned to males in half the mixed-sex groups and to females in the other half.

Other Measures. Subjects were asked about their satisfaction with the solution reached by their group and with their amount of influence over the solution, using Likert-type six-point questions. Values from 1 to 6, indicating increasing degrees of satisfaction were assigned to the alternatives: very dissatisfied, somewhat dissatisfied, neither satisfied nor dissatisfied, fairly satisfied, quite satisfied, very satisfied.

Subjects

Three hundred and eighty-four subjects, half of each sex, were recruited from various freshman and sophomore level courses at the University of Michigan. The problem was multiple role played (Maier & Zerfoss, 1952) in 10 sessions.

RESULTS

Conflict Conditions

The principal hypothesis of this study is that an increase in the strength of the workers' feelings about the monotony of their jobs causes them to oppose effectively the foreman's suggested change and that the resulting conflict generates a higher proportion of integrative solutions. Table 1 presents the proportions of solutions of each type produced by groups in each conflict condition, as well as separately by the type of foreman and type of workers. A comparison of the results for groups with Strong as against Weak workers supports the principal hypothesis at the .01 level of confidence by t test. Integrative solutions were produced by 45.8% of the groups with Strong worker roles and by only 18.8% of the groups with Weak worker roles, disregarding the type of foreman's role. The resistance of the Strong workers to the foreman's favored solution (new) is apparent from the table. Less than half as many new solutions were accepted by groups with Strong workers as by groups with Weak workers. This suggests that integrative solutions resulted when the point of view opposing the foreman's was strongly held and a conflict developed which forced the groups to create different types of solutions as a way of reaching agreement.

On the other hand, Table 1 also shows that altering the foreman's role had little effect. Only a slightly higher proportion of integrative solutions was produced by groups with Considerate foreman's roles (35.4%) as compared with groups with Dominant foreman's roles (29.2%), disregarding the type of workers' roles. Groups accepted new solutions more frequently than either old or integrative solutions under both types of foremen, although the Dominant foremen were somewhat more successful in having their way.

When the workers had weak feelings about the monotony of the jobs, even the Considerate foremen seem to have dominated the discussion, as evidenced by the fact that the same high proportion of new solutions (66.7%) was produced by the Dominant and Considerate foremen groups in the Weak workers condition. Where the workers' feelings were strong, on the other hand, the Considerate foremen appear to have been somewhat more responsive to them than the Dominant foremen were (Table 1).

TABLE 1

DISTRIBUTIONS OF SOLUTIONS BY GROUPS IN EACH CONFLICT CONDITION[1]

Type of solution	Type of group			
	Strong workers	Weak workers	Dominant foreman	Considerate foreman
	(N = 48)	(N = 48)	(N = 48)	(N = 48)
Integrative[a]	45.8%	18.8%	29.2%	35.4%
New	31.2	66.7	54.2	43.8
Old	22.9	14.6	16.7	20.8
	99.9%	100.1%	100.1%	100.0%
Foreman: Workers:	Dominant Strong	Considerate Strong	Dominant Weak	Considerate Weak
	(N = 24)	(N = 24)	(N = 24)	(N = 24)
Integrative[a,b]	37.5%	54.2%	20.8%	16.7%
New[b]	41.7	20.8	66.7	66.7
Old[b]	20.8	25.0	12.5	16.7
	100.0%	100.0%	100.0%	100.1%

[a] Tests of differences in percentage of integrative solutions produced by groups with: Strong vs. Weak workers—$z = 2.93$, $p < .01$; Dominant vs. Considerate foremen—$z = 0.65$, ns; Strong workers with Dominant foreman vs. Considerate foremen—$z = 1.15$, ns.

[b] For relationship among conflict conditions and type of solution: chi square = 11.30, $.10 > p > .05$.

TABLE 2

SOLUTIONS IN GROUPS OF DIFFERENT SEX COMPOSITION

Type of solution	Type of group							
	Mixed sex		Single sex[a]		All male		All female	
	N	%	*N*	%	*N*	%	*N*	%
Integrative[b]	13	40.6	18	28.1	12	37.5	6	18.8
New	10	31.2	37	57.8	17	53.1	20	62.5
Old	9	28.1	9	14.1	3	9.4	6	18.8

[a] Single sex combines the results of the all-male and all-female groups.
[b] Test of difference in percentage of integrative solutions produced by mixed-sex and single-sex groups, $z = 1.19$, *ns*.

Sex Composition

Table 2 shows the distributions of solutions produced by groups of each sex composition, regardless of the conflict conditions. While a z test of the difference in the proportions of integrative solutions produced by mixed-sex groups and single-sex groups is not significant ($z = 1.19$, $p = .15$), certain differences are apparent. The mixed-sex groups produced the lowest proportion of new solutions, which raised their proportions of both integrative and old solutions.[2] This result parallels in more modest fashion the effect produced by strengthening the workers' commitment to the old method (see Table 1). Also, the lowest proportion of integrative solutions was produced by the all-female groups.

Interaction of Conflict Conditions and Sex Composition

An examination of the combined effects of the conflict conditions and the sex composition of the groups indicates that the trend shown in Table 2 was even more marked in the strong worker condition. The mixed-sex groups in the Strong worker condition produced the smallest proportion of new solutions and these were distributed between integrative and old solutions (Table 3). However, an insufficient number of the mixed-sex, Strong worker groups produced integrative solutions, as compared with the number produced by single-sex, Strong worker groups, to confirm statistically the hypothesis that these two variables would interact to produce a substantially higher proportion of creative solutions. The proportions of integrative solutions produced by these two types of groups were not statistically significant ($z = 1.15$, $p = .15$).

A chi square test of the over-all relationship among sex composition, worker commitment, and type of solution is statistically significant (chi square = 15.58, $p < .05$). The significance arises from the small proportion of new solutions (6.2%) accepted by the mixed-sex, Strong worker groups, with an accompanying increase in the production of integrative and old solutions, as compared to the high proportion of new solutions (71.9%) obtained from the single-sex, Weak worker groups. Moreover, the linear increase in the proportion of integrative solutions produced as the conflict and sex composition factors are added offers suggestive support for the hypothesis concerning the interaction of these variables. The differences in proportions of integrative solutions produced by the Weak worker, single-sex groups and both the Strong worker, mixed-sex and Strong worker, single-sex groups are significant at the .01 level of confidence, while the differences between these groups and the Weak worker, mixed-sex groups were not significant (z tests).

One other point of interest in Table 3 is that the six integrative solutions produced by the all-female groups all occurred in the Strong worker conditions.

Satisfaction with Influence and with the Solution

The basic assumption underlying this experiment is that the free interplay of the different ideas present in a group is fundamental to the development of high quality solutions to problems. The attempt to increase the workers' commitment to their present work method was designed to improve the quality of the solutions by permitting the workers' views to be placed in effective opposition to the foreman's, and, thus force their consideration by the group. The advantages and disadvantages of both the foreman's and the workers' viewpoints can thereby be discov-

[2] The distributions of solutions in mixed-sex groups led by male foremen were not different from those led by female foremen, so the results will be reported together.

TABLE 3

SOLUTIONS BY STRONG AND WEAK WORKER GROUPS OF DIFFERENT SEX COMPOSITION

Type of solution[a]	Type of group															
	Strong workers								Weak workers							
	Mixed sex		Single sex		All male		All female		Mixed sex		Single sex		All male		All female	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Integrative[b]	9	56.2	13	40.6	(7)	(43.8)	(6)	(37.5)	4	25.0	5	15.6	(5)	(31.2)	(0)	(0.0)
New	1	6.2	14	43.8	(6)	(37.5)	(8)	(50.0)	9	56.2	23	71.9	(11)	(68.8)	(12)	(75.0)
Old	6	37.5	5	15.6	(3)	(18.8)	(2)	(12.5)	3	18.8	4	12.5	(0)	(0.0)	(4)	(25.0)

[a] For relationship among worker commitment, sex composition (mixed vs. single), and type of solution: chi square = 15.58, $p < .05$. The numbers in parentheses were not included in this calculation.
[b] Test of difference in percentage of integrative solutions produced by Strong worker groups of mixed- vs. single-sex, $z = 1.00$, *ns*.

ered and a joint effort can be made to create a new solution, which will incorporate the advantages and minimize the disadvantages. One measure of the extent to which the problem solving process was a joint one, rather than unilateral, would be that both the foreman and the workers were satisfied with the amount of influence they had over the solution.

Groups were classified as being mutually satisfied with their influence over the solution, according to the subjects' responses to the question about their satisfaction with influence, where the foreman *and* all the workers checked responses 5, quite satisfied, or 6, very satisfied. According to this stringent criterion, 12 of the 96 groups were mutually satisfied with their influence over the solution. Table 4 shows that 8 of these 12 groups (66.7%) produced integrative solutions, compared to only 23 of the 84 groups not mutually satisfied with their influence (27.7%). This difference is significant at the .05 level of confidence. Interestingly, 6 of the 12 groups (50.0%) mutually satisfied with their influence over the solution occurred in the Considerate foreman-Strong worker condition, and all 6 produced integrative solutions. It is apparent that the Considerate foremen did not capitulate to the Strong workers' feelings and often successfully initiated change through a process of joint problem solving.

The method of increasing worker influence used in this experiment did not necessarily ensure acceptance of the integrative solutions thereby produced, however. Table 5 shows the percentage of groups in which the foremen and workers, respectively, were satisfied with each type of solution and in which they were mutually satisfied with the solution. It will be seen that the foremen were least often satisfied with the old solutions ($p < .05$, z test) and the workers were least often satisfied with the new solutions ($p < .05$, z test). The proportions of foremen and workers satisfied with the integrative solutions were no greater than the proportions satisfied with the solution they preferred initially. However, the same proportion of foremen was satisfied with the integrative solutions as with the new ones, and only a somewhat higher proportion of workers was satisfied with the old solutions than with the integrative.

TABLE 4

SOLUTIONS AND SATISFACTION WITH INFLUENCE OVER THE SOLUTION

Type of group	Type of solution					
	Old		New		Integrative[a]	
	N	%	*N*	%	*N*	%
Mutually satisfied with influence over solution[a]	1	8.3	3	25.0	8	66.7
Not mutually satisfied with influence over solution	16	19.3	44	53.0	23	27.7

[a] Groups were classified as mutually satisfied with influence when the foreman *and* all the workers responded 5, quite satisfied, or 6, very satisfied. One group could not be classified because its foreman did not indicate his satisfaction with influence over an old solution.
[b] Percentage difference significant at the .05 level of confidence, $z = 2.11$.

TABLE 5

TYPE OF SOLUTION AND SATISFACTION WITH THE SOLUTION

	Type of solution					
	Old		New		Integrative	
	N	%	*N*	%	*N*	%
Foremen[a]						
Satisfied[b]	3	17.6[c]	30	63.8	20	64.5
Not satisfied	14	82.4	17	36.2	11	35.5
Workers						
Satisfied[b]	41	75.9	70	49.6[c]	55	59.1
Not satisfied	13	24.1	71	50.4	38	40.9
Satisfaction in groups[a]						
Mutually satisfied[b]	2	11.8	4	8.5	5	16.1
Not mutually satisfied	15	88.2	43	91.5	26	83.9

[a] One foreman did not indicate his satisfaction with an old solution.

[b] Responses of 5, quite satisfied, and 6, very satisfied, were classified as satisfied; all others as not satisfied. Groups were classified as mutually satisfied where the foreman *and* all the workers were satisfied.

[c] Percentage satisfied is smaller for this solution than for both others at the .01 level of confidence.

The foremen and workers were somewhat more likely to be mutually satisfied with the integrative solutions than with either the old or new solutions. Also, four of the five groups mutually satisfied with their integrative solutions were also mutually satisfied with their influence over the solution, while only one of the six groups mutually satisfied with the old or new solutions were also mutually satisfied with their influence. These results suggest that where the foreman and workers freely exchanged ideas in solving the problem, both the quality and acceptance of the solutions were improved.

DISCUSSION

This experiment was designed to test the hypothesis that conflict of ideas causes groups to search for alternatives and thereby to improve the quality of their solutions. The results support this hypothesis. Increasing the workers' resistance in this situation, where the foreman usually dominates the decision, resulted in an increased proportion of integrative solutions. On the other hand, when the workers did not feel strongly about the monotony of their jobs, they submitted fairly readily to the foreman's request to adopt the new work method and the groups rarely developed integrative solutions.

The data suggest, however, that if high acceptance as well as high quality of the solutions is to be achieved, the conflict of ideas must occur within an atmosphere of tolerance for others' points of view. Where such an atmosphere was lacking, the conflict of ideas which occurred when the workers held strong feelings often produced integrative solutions which lacked strong member satisfaction with the outcome. The struggle involved in arriving at a solution may have caused hard feelings in the group and have lowered the group's enthusiasm for the agreed-upon work method.

The lack of strong over-all effect of making the foreman Considerate rather than Dominant, especially where the workers' feelings were Weak, reinforces the importance of conflict for achieving high quality solutions. The Weak workers offered little opposition to the foreman's proposition even where he was willing, supposedly, to listen to their views. Without strong opposition to the new method, there was no cause for the group to consider possible alternative solutions. The foreman was able to convince the workers of the benefits of the new method, and by virtue of his authority over them, get them to accept it on a trial basis even though they were not satisfied with the method. To achieve creative solutions to problems under these conditions, a superior should probably encourage the expression and consideration of opposing points of view, rather than try to persuade the group of the merits of his own ideas. There was some suggestion in the data from groups with Strong workers that the Considerate foremen were more likely to achieve integrative solutions than were the Dominant foremen, possibly because they were more prone to consider seriously the opposition point of view.

The interactive effects on problem solving of the sex composition of the groups and the conflict conditions deserve comment. Since women are generally poorer problem solvers than men, one might have expected them to reduce the effectiveness of the males in the mixed-sex groups. Thus the proportion of integrative solutions from the mixed-sex groups

might have been expected to have been approximately the mean proportion produced by the all-male and all-female groups. Instead, the proportion of integrative solutions produced by the mixed-sex groups was slightly greater than that produced by the all-male groups, especially in the Strong worker condition, and substantially greater than the proportion produced by the all-female groups.

These results provide further suggestive evidence to support the proposition that part of women's usual poorer problem solving performance is due to a lack of motivation related to their sex-role identification (Milton, 1957, 1959). Hoffman and Maier (1961b) had shown previously that mixed-sex groups seemed to stimulate their female members to greater and more effective effort. In the present study, both the slightly improved performance of the mixed-sex groups in the Strong worker condition and the fact that all the integrative solutions produced by all-female groups were produced in the Strong worker condition, suggest that these role instructions may have produced a higher degree of involvement in the problem among the female subjects.

The general findings of this study seem to support the following conclusions:

1. Different perspectives lead to different "directions" for identifying a problem (Maier, 1930) and to alternative possibilities for solutions and, therefore, offer the potential for creative problem solving.

2. Different perspectives, when set in effective opposition to each other, can delay the premature acceptance of an obvious solution and contribute to turning a choice situation into a problem. A group may then be forced to search for alternative solutions which might better satisfy the requirements of the problem.

3. The joint resolution of such conflict by all members of the group leads to solutions of high quality which are well accepted by the group members.

SUMMARY

This experiment was designed to test the following hypothesis: conflict generated by increasing the resistance of subordinate members of a group to a solution offered by a member in an authority position increases the frequency of high quality solutions. The supplementary hypothesis that mixed-sex composition in interaction with such conflict further enhances the production of high quality solutions was also tested.

Ninety-six four-person groups role played the Change of Work Procedure problem under conditions where a foreman was either dominant or considerate in combination with subordinate workers whose commitment to the *status quo* was either relatively strong or weak. Eight groups each of all-male, all-female, and mixed-sex composition were assigned at random to each condition.

The results support the major hypothesis. High quality solutions were produced by 45.8% of groups under Strong worker conditions and by only 18.8% of groups under Weak worker conditions. The hypotheses concerning the effects of sex composition were only supported suggestively.

The discussion emphasizes the importance for achieving problem solutions of both high quality and acceptance, of encouraging the expression and consideration of conflicting points of view in an atmosphere of mutual influence among the group members. Differences of opinion can promote creative problem solving by turning a choice situation into a problem and by providing a variety of directions in thinking.

REFERENCES

GHISELLI, E. E., & LODAHL, T. M. Patterns of managerial traits and group effectiveness. *J. abnorm. soc. Psychol.*, 1958, **57**, 61–66.

HOFFMAN, L. R. Homogeneity of personality and its effect on group problem-solving. *J. abnorm. soc. Psychol.*, 1959, **58**, 27–32.

HOFFMAN, L. R., & MAIER, N. R. F. Quality and acceptance of problem solutions by members of homogeneous and heterogeneous groups. *J. abnorm. soc. Psychol.*, 1961, **62**, 401–407. (a)

HOFFMAN, L. R., & MAIER, N. R. F. Sex differences, sex composition, and group problem solving. *J. abnorm. soc. Psychol.*, 1961, **63**, 453–456. (b)

MAIER, N. R. F. Reasoning in humans: I. On direction. *J. comp. Psychol.*, 1930, **10**, 115–143.

MAIER, N. R. F. *Principles of human relations*. New York: Wiley, 1952.

MAIER, N. R. F. An experimental test of the effect of training on discussion leadership. *Hum. Relat.*, 1953, **6**, 161–173.

MAIER, N. R. F. *The appraisal interview: Objectives, methods and skills.* New York: Wiley, 1958.

MAIER, N. R. F., & HOFFMAN, L. R. Quality of first and second solutions in group problem solving. *J. appl. Psychol.*, 1960, **44**, 278–283.

MAIER, N. R. F., & HOFFMAN, L. R. Organization and creative problem solving. *J. appl. Psychol.*, 1961, **45**, 277–280.

MAIER, N. R. F., & ZERFOSS, L. F. MRP: A technique for training large groups of supervisors and its potential use in social research. *Hum. Relat.*, 1952, **5**, 177–186.

MILTON, G. A. The effects of sex-role identification upon problem solving skill. *J. abnorm. soc. Psychol.*, 1957, **55**, 219–244.

MILTON, G. A. Sex differences in problem solving as a function of role appropriateness of the problem content. *Psychol. Rep.*, 1959, **5**, 705–708.

TORRANCE, E. P. Some consequences of power differences on decision making in permanent and temporary three-man groups. In A. P. Hare, E. F. Borgatta, & R. F. Bales (Eds.), *Small groups.* New York: Knopf, 1955. Pp. 482–492.

(Received February 14, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 64, No. 3, 215–221

PERCEPTUAL RECOGNITION AS A FUNCTION OF ANXIETY, MOTIVATION, AND THE TESTING SITUATION [1]

E. D. LONGENECKER [2]

Pennsylvania State University

Common to most studies investigating the influence of anxiety on behavior is the fundamental hypothesis that anxiety acts primarily as a drive stimulus. In the work of Spence and Taylor and their associates involving simple, one response situations such as conditioning there appears to be a direct relationship between anxiety and performance level. Their assumption is that anxiety level reflects a difference in the subject's drive which combines with the habit strength of the dominant response in the subject's response hierarchy to raise and strengthen the excitatory potential. On the other hand, in more complex problem solving situations where no single response dominates the response hierarchy, anxiety level is found to be inversely related to performance. The assumption here is that the higher anxiety level increases the response strength for all responses in the response hierarchy. Since the strength of competing responses will be raised accordingly, the probability of the occurrence of the correct response will be lowered. Put another way, the more competing responses brought to a threshold level, as one's drive level is increased, the more interference there will be with the completion of the problem solving task.

Mandler and Sarason (1952) approach the problem somewhat differently. They postulate that two fundamental drives are activated simultaneously in any problem solving situation. First, there is an anxiety drive which is capable of eliciting all responses in the individual's response hierarchy regardless of their relevance to the completion of the task. It is assumed that the task relevant responses are specific to each task and that their habit strength is built up only as a result of the reinforcement gained upon successful completion of each item on the task. In addition, there are the task irrelevant responses which are assumed to be competing responses existing in the individual's hierarchy of responses as a result of past experiences. These competing, task irrelevant responses are generalized to each new task and are assumed to hinder successful performance.

Thus it is postulated that, at the beginning of a serially presented task when the strength of the relevant responses is low, anxiety functions as a negative drive. As the task relevant responses are reinforced and the generalized irrelevant responses are not reinforced, however, anxiety gradually comes to function as a positive drive. In a series of timed, perceptual recognition items it would be predicted that high anxious (HA) subjects would have significantly slower reaction times on the early trials than would low anxious (LA) subjects. As the succeeding items of the tests are successfully completed, however, the reaction times of the HA subjects would gradually approach and eventually become significantly faster than those of the LA subjects.

A second drive which Mandler and Sarason (1952) believe to be operating in a testing situation they call the "task drive," perhaps akin to achievement motivation. According to these authors, this drive elicits only the so-called task relevant responses, or those responses leading to successful completion of the task. Thus a direct relationship between the strength of the task drive and performance on a given task is postulated. Consequently, in our recognition tests it would be predicted that subjects with a high task drive (i.e., high motivation, HM), would have significantly faster reaction times than would

[1] This study is based on a dissertation submitted in partial fulfillment of the requirements for the PhD degree at the University of Texas, 1956.

[2] The author wishes to express his sincere appreciation for the encouragement and direction given by his committee chairman, Wayne H. Holtzman, throughout the study.

subjects with a low task drive (i.e., low motivation, LM).[3]

Although Mandler and Sarason (1952) state that anxiety and motivation (i.e., their task drive) are influenced by the nature of the testing situation, instructions, etc., they do not state precisely how these variables are expected to influence these drives. On the assumption that thresholds for the arousal of both anxiety and motivation are lower for subjects who are high on these variables than they are for subjects who are low on these variables, it may be expected that ego involving conditions would have a differential effect on both the subjects high or low in anxiety and the subjects high or low in motivation. To illustrate: stress conditions will heighten the functional level of anxiety in the HA subjects but will not heighten the functional level of the anxiety drive in the LA subjects. Thus it would be predicted that HA subjects given a series of tests under stress conditions would have significantly slower reaction times on the early trials of each test than HA subjects given the tests under nonstress conditions. On subsequent trials of each test, this significant difference would eventually reverse itself. In contrast, LA subjects given the tests under stress conditions would have reaction times that are not significantly different from LA subjects given the tests under nonstress conditions.

Applying the same assumption to the variable of motivation, stress conditions will heighten the level of motivation of the HM subjects but will not heighten the level of motivation of the LM subjects. Thus it would be predicted that HM subjects given the tests under stress conditions would have reaction times that are significantly faster than for the HM subjects given the tests under nonstress conditions. However, the LM subjects given the tests under stress conditions would have reaction times that would not be significantly different from LM subjects given the tests under nonstress conditions.

The following study is designed to investigate these conceptualizations of the functions of anxiety, a task oriented drive (i.e., motivation), and stress.

[3] For simplicity's sake, only the term Motivation will be used in referring to this variable throughout this paper.

METHOD

Experimental Design and Subjects

The basic design is a 3×2 analysis of variance with the variables of Anxiety and Motivation being dichotomized into high and low groups while the third variable represents Stress and Nonstress testing conditions. All subjects were sophomore male students from the University of Texas with ACE scores within the range from 90 to 120 inclusive (approximately the middle 50% of the students admitted to the university) who met the criteria defining the variables of Anxiety and Motivation. The experimental control of sex and intelligence was necessitated by the significant correlations reported by several authors (Mandler & Cowen, 1958; Rockett, 1955; Sarason, 1957a, 1957b, 1959; Sarason & Mandler, 1952; Thurstone, 1944; Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954) between these variables and the criterion measures and the perceptual recognition tests used in this study.

Procedure

Measures of anxiety and motivation. HA and LA subjects were selected on the basis of their extreme scores on the Sarason Scale of Test Anxiety. The cutoff scores used to select these two groups represented the 28th and the 70th percentile points on the distribution of scores for a randomly selected group of 292 male subjects with ACE scores as defined above. The main reason for preferring this measure of anxiety over the other measures now available is based on the fact that any measure of "anxiety" only measures what might be called "anxiety proneness" to certain stimulus situations. Thus it would seem important that the specific stimulus situation used in the experiment to arouse anxiety be closely related to the criterion measure.

There were several other reasons for using this scale. First, this scale has been standardized on a college population, and Gordon and Sarason (1955) report split-half and test-retest reliabilities of .91 and .82, respectively. Furthermore, its disguise as a measure of test anxiety (it is administered as an attitude questionnaire) and its ability to elicit the interest of the subjects completing it seemed most desirable.

HM and LM subjects were selected on the basis of a combination of under- and overachievement in college grades and low and high need Achievement scores on the Edwards Personal Preference Schedule (PPS). In selecting subjects with records of either under- or overachievement, students having either a grade point ratio (GPR) of 1.1 or less (i.e., underachievers) or a GPR of 1.9 or greater (i.e., overachievers) were selected with GPR being measured on a three-point scale. On the need Achievement scale of the Edwards PPS, the second measure used in defining motivation, cutoff scores of 13 and 18 were selected. These cutoff scores represented the 35th and the 72nd percentile points, respectively, on the distribution of scores for the 292 subjects completing the criterion

measures. Thus for a subject to be included in the LM group, he had to have a low GPR *and* a low score on the need Achievement scale; and for a subject to be included in the HM group he had to have a high GPR *and* a high score on the need Achievement scale.

The reason for using this dual criterion measure for motivation stems primarily from the distinction between the concepts of "need" and "drive." It is conceivable that certain subjects will have a high "need" for achievement but will be unable, for some reason or other, to exert themselves over a long period of time such as is demanded in maintaining a GPR up to their levels of capacity. For the purposes of the present study, it seemed important to obtain students who not only indicated a high or low level of aspiration for achievement as is most likely to be shown by the Edwards PPS score but also to have some evidence that they have shown a "drive" level which has been consistent with the "need" level. Thus it seemed wise to select only those subjects scoring either high-high or low-low in terms of their "expressed need" to achieve and their actual achievement.

Relationships of criterion measures. Since four different measures were obtained in defining the criterion variable of Anxiety and Motivation, it was important to know if they were related in any way. The correlations between these measures are, therefore, presented in Table 1. The results show that test anxiety is unrelated to any of the other measures, indicating the statistical independence of the measures defining Anxiety and Motivation. The correlations between the ACE and GPR and anxiety are also not significant. Nonsignificant correlations between ACE scores and perceptual recognition tests also demonstrate that the variable of intelligence was successfully partialed out by the experimental procedures.

The only correlation reaching statistical significance is between the ACE and need Achievement. This reflects the fact that one's need for achievement influences performance on the ACE to a small but significant degree.

Testing conditions. By defining the criterion groups according to the selected measures, it was possible to obtain 72 subjects with 18 subjects in each of the four criterion groups (i.e., HA-HM, LA-LM, HA-LM, and LA-HM). Half of the subjects in each of these groups were randomly selected to take the tests under Stress conditions while the other subjects were given the tests under Nonstress conditions.

There were two basic variations in the testing situation under the Stress and Nonstress conditions. These variations were in the manner of conducting the testing, and the nature of the instructions given to the subjects.

The manner of conducting the testing session varied from one of formality under the Stress conditions to one of informality under the Nonstress conditions. In the Stress situation the experimenter was always dressed in a business suit, and he always introduced himself and addressed the subject by using the title "Mister," followed by the appropriate surname. In general, it could be assumed that this situation was most characteristic of a student-instructor relationship.

TABLE 1

CORRELATIONS BETWEEN MEASURES OBTAINED IN DEFINING CRITERION VARIABLES

	ACE	GPR[a]	Need Achievement	Anxiety
ACE				
GPR[a]	.03			
Need Achievement	.23*	.07		
Anxiety	.00	.04	−.04	

[a] Correlations with GPR are biserial corrected for widespread groups, others are product-moment. All correlations are from 292 sophomore, male students at the University of Texas with ACE scores and GPRs as defined in text.
* Significant beyond .001 level of confidence.

Throughout the Nonstress situation the atmosphere fostered by the experimenter was more characteristic of a student-student relationship. Consequently, in this situation the experimenter always dressed in sportswear, always introduced himself by using only his first and last names, and always used the subject's first name when giving instructions and answering questions.

In addition to varying the general atmosphere of the testing situation, the instructions given the subjects in the two situations varied in a number of ways so as to act consistently with the changes in the atmosphere. For instance, subjects taking the tests under the Stress conditions were told that they were about to take a "new type of intelligence test" that was known to be highly related to grades and academic standing; they were kept in the dark as to the nature or composition of the test; and they were told that this test has so far proven to be quite easy for college students. In the Nonstress situation the subjects were told that the test they were about to take was relatively new and, therefore, "nobody knew what it measured, to what it might be related, or the meaning of the results." Furthermore, they were given a general description of the different parts of the test in addition to being told that some parts of the test were quite difficult and that no one was expected to be able to give the correct answers to all of the cards.

In setting up these variations in the testing situation, it was realized that it would be impossible to know which variables in the situations brought about a change in performance on the tests, if any changes were found at all. Such an analysis was not, however, the object of study. Instead, the primary purpose was to try to study the effect, if any, of changing the degree of threat or ego involvement for the subject as it may be related to the variables of Anxiety and Motivation.

Perceptual recognition measures. All subjects meeting the criteria defining the HA or LA and the HM or LM groups were individually given five perceptual recognition tests in the following order: Holtzman Figure-Recognition Test, achromatic form; Holtzman Figure-Recognition Test, chromatic form;

TABLE 2

ANALYSIS OF VARIANCE FOR THE HOLTZMAN ACHROMATIC FIGURES

Source of variation	*SS*	*df*	*MS*	*F*
Anxiety (A)	748	1	748	
Motivation (M)	1,240	1	1,240	
Testing conditions (TC)	866	1	866	
A × M	6,613	1	6,613	
A × TC	122,584	1	122,584	12.36*
M × TC	94,033	1	94,033	9.38*
A × M × TC	589	1	589	
Between-subjects, within-groups	637,193	64	9,958	
Total between-subjects	863,867	71		
Between trial blocks (TB)	481,511	3	160,504	56.46*
TB × A	7,896	3	2,632	
TB × M	1,323	3	441	
TB × TC	5,712	3	1,904	
TB × Pooled conditions	30,045	12	2,504	
TB × Pooled subjects within groups	545,839	192	2,843	
Total	1,936,193	287		

* Significant at .001 level of confidence.

Witkin's Embedded Figures Test; Hidden Digits Test; and Mooney Closure Test. All of these tests are similar in structure and in design. Essentially they are tests measuring skill in recognition of a definite animate or inanimate object "hidden" among an ambiguous configuration of perceptual cues that make up the background stimuli. The Holtzman tests are made up of animate figures hidden among a configuration of either chromatic or achromatic background stimulus cues. The other tests are described in detail by Thurstone (1944).

Reaction time scores were taken on all items of each test. An inverse square root transformation of these scores was used in the statistical analysis on only four of the tests, however. On the Mooney Closure Test, where a number of the items were incorrectly identified within a 2-minute time limit, it was necessary to use a pass-fail criterion for scoring and a square root transformation in the statistical analysis.

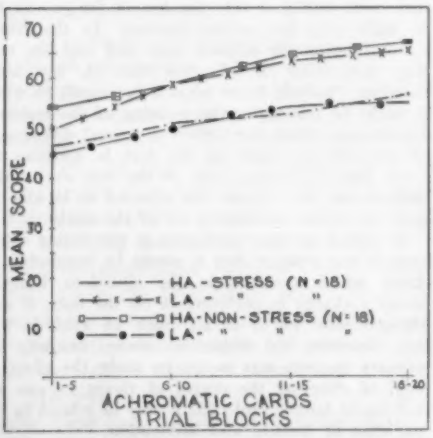

FIG. 1. Performance curves on the Holtzman Achromatic Cards for HA and LA groups under Stress and Nonstress conditions. (Curves are based on five-trial averages for each group. Additional markings are to aid in identifying each specific curve.)

From empirical studies, especially those done by Thurstone (1944), it is possible to gain some understanding of what is being measured by these tests. In his factorial study of perceptual tests, Thurstone found the original embedded figures, the hidden digits, and the original street-gestalt version of the Mooney Figures to have a common factor which he thought was measuring the ability to form a perceptual closure against distraction. In addition to having loadings on this factor, the Gottschaldt figures and the hidden digits were shown to have high loadings on a second factor which Thurstone thought represented flexibility in dealing with conflicting gestalts. These factors are commonly acknowledged as part of the spatial relations factors found on most intelligence tests.

Although the Holtzman Figures have not previously been correlated with other perceptual tests, they were constructed to measure perceptual closure against distracting and conflicting stimuli. Therefore, it would be expected that perceptual processes similar to those identified in the other tests in this battery are operating in this test. The correlations obtained in the process of carrying out this study support this conclusion. The intercorrelations of all these tests range from .41 to .70, all significant at the .001 level of confidence or better.

RESULTS

The results of an analysis of variance for the Holtzman recognition test, achromatic

version, are presented in Table 2. The findings for this test are exactly the same as the other tests with the one exception of the Hidden Digits Test where the *F* ratios for A × TC and M × TC only approach statistical significance at the .05 level of confidence. A low odd-even reliability coefficient of .58 for this test would seem to account for these lowered *F* ratios.

The performance curves presented in Figure 1 illustrate the specific interpretations that can be derived from the analysis of variance for the anxiety groups. Curves for the HM and LM groups are essentially replications of those shown in Figure 1. Thus we find that both Anxiety and Motivation, as defined by the criterion measures, function in the same way in each of the testing situations.

In Figure 1 the curves for the HA Stress group and the LA Nonstress group are not significantly different from each other. Both curves reflect a gradual but significant improvement in performance from trial to trial. Likewise, the curves for the LA Stress group and the HA Nonstress group are not significantly different from each other. Again both curves reflect a steady improvement in performance. Thus all groups show an improved performance across trials that is significant. However, there is no significant interaction between this practice effect and the variables of Anxiety, Motivation, or the testing conditions.

There are, however, very significant interactions between Anxiety and the testing conditions and between Motivation and the testing conditions. Furthermore, these interactions are in the same direction. Thus the LA Stress and the HA Nonstress groups have shorter reaction times than do the LA Nonstress and the HA Stress groups.[4] Likewise, the LM Stress and the HM Nonstress groups have shorter reaction times than do the LM Nonstress and the HM Stress groups. *Both* the LA and the LM groups have shorter reaction times than their respective HA and HM groups under the Stress conditions, while *both* the HA and the HM groups have shorter reaction times than their respective LA and LM groups under the Nonstress conditions.

[4] The use of inverse square root transformations accounts for the reversed appearance of the curves in Figure 1.

DISCUSSION

Anxiety

The significantly longer reaction times of the HA group over the LA group in the Stress situation are consistent with the belief that anxiety functions as a negative drive stimulus which impedes performance on a problem solving task. However, the fact that this difference remains from trial to trial is inconsistent with our predictions and with the results of Mandler and Sarason (1952) where they found that early differences between the HA and the LA groups disappeared gradually as reinforcement was obtained from trial to trial. These results appear to be contradictory to their original ideas. However, closer examination of the basic conditions underlying the interaction of Trial × Anxiety found by Mandler and Sarason reveals that what these authors consider a Trial × Anxiety interaction is probably a Situation × Anxiety interaction. Thus during the early trials of their task, the situation is more threatening to the HA group than it is to the LA group; however, during the later trials the stressful aspects of the situation disappear with the subsequent disappearance of the difference in the levels of performance of the two anxiety groups.

In the present study, it seems that the stress or nonstress aspects of the testing conditions remained relatively constant throughout each test as a result of the deliberate attempt to create *and* maintain either a competitive and threatening atmosphere or a noncompetitive and relaxed atmosphere. However, in the Mandler and Sarason (1952) study, the testing conditions were designed to be more neutral; and once the threat presented by the novelty of their situation subsided, the improved performance was manifested by the HA group.

Although the inferior performance of the HA over the LA subjects in the Stress situation is as predicted, the fact that the HA group has significantly faster reaction times than the LA groups when the tests are given in the Nonstress situation represents a complete reversal of our prediction for the Nonstress conditions. This is in sharp contrast to the results in the Stress situation where some differential, negative drive force is operating.

These positive and negative drive aspects of anxiety are not basically unlike those originally postulated by Mandler and Sarason (1952). Instead of finding anxiety to change functional characteristics across trials, it is necessary to consider anxiety as having different functional characteristics according to the Stress, Nonstress features of the testing situation.

Two studies by Sarason (1956, 1957a) have demonstrated similar functional characteristics of anxiety. Using college students he found that "motivating instructions" detrimentally affected the performance of HA subjects while LA and MA subjects were facilitated on a serial learning task. In another very similar experiment testing the retention of nonsense syllables, it was found that HA subjects were significantly superior to LA subjects given neutral instructions while the opposite results were obtained under so-called subject oriented instructions emphasizing "the need for *S* to perform well in order to maintain his self-esteem" (Sarason, 1957a, p. 171).

In another study by Ruebush (1960) on sixth grade boys with Otis beta IQ scores between 111 and 122, it was found that the HA subjects were inferior to LA subjects on a short form of the Witkin's Embedded Figures Test (Jackson, 1956). In analyzing each subject's approach to the testing situation by means of a cautiousness index, the author concludes that the differential "effect of anxiety on performance . . . is mediated primarily by defensive reactions to the anxiety"—and that these effects "may vary systematically depending upon both intelligence level and type of task and instruction" (Ruebush, 1960, p. 211).

These studies, along with the present results, indicate that anxiety can function *both* as a positive or as a negative drive on the same task depending on the stress, nonstress characteristics of the situation in which the tasks are completed. This being the case, it would seem that anxiety must be considered not only in terms of the type of response involved, but also in terms of the *task situation*.

Some explanation must be found for the "reversal effect," namely, that HA subjects do more poorly under stress than under lack of stress, whereas the LA subjects do better under stress than under lack of stress. One interpretation is that, up to a moderate level of intensity, anxiety is directly related to the level of performance; however, once past this threshold, anxiety is inversely related to level of performance. Under such conditions, it would be expected that when anxiety is at a medium level subjects would perform at their highest level while it would be expected that under conditions where anxiety is at a very low intensity or is at a very high intensity, subjects would perform at their lowest level. In that the LA group under Stress and the HA group under Nonstress performed significantly better than did the LA group under Nonstress and the HA group under Stress, these expectations are consistent with the results of this study.

The results from several studies reported by Matarazzo and his colleagues are also consistent with this interpretation. In two studies using college students (Matarazzo & Phillips, 1955; Matarazzo, Ulett, & Saslow, 1955), one involving maze learning, the other using a digit symbol test, the MA groups performed on a consistently higher level than the HA and LA groups. In a third study using Veterans Administration inpatients on a pursuit-metor task (Matarazzo & Matarazzo, 1956), similar trends were observed.

The concept of "activation" or "energy mobilization" as developed by Duffy (1957) seems to be applicable to this interpretation. This concept "is designed to describe the intensity aspect of *all* behavior" and is defined as "the extent to which the organism as a whole is activated" (p. 266). Although this concept refers to the degree of internal arousal, Duffy acknowledges the possibility of a discrepancy between the degree of arousal and the overt response, a discrepancy which she believes to be due to the intervention of inhibitory processes. These inhibitory processes could be conceived of in terms of the competing or irrelevant responses of Spence and Sarason. Furthermore, Duffy also believes that the degree of optimal activation appears to be a moderate one, the curve expressing the relationship between activation and quality of performance taking the form of the inverted **U**. Thus the disorganization of responses during overstimulation (e.g., our

HA subjects under stress) would represent one end of this U shaped curve.

Motivation

Since Mandler and Sarason (1952) tried to integrate their ideas about the functioning of anxiety with those of a task oriented drive, an attempt was made to define this variable and predict its functional characteristics. Thus it was predicted that HM subjects would have significantly faster reaction times than LM subjects. This prediction is supported only under Nonstress conditions. Under Stress conditions, the same "reversal effect" appears, since the LM group is found to have significantly faster reaction times than the HM group. Thus the results testing the effects of stress on this drive variable are identical to those of anxiety.

In the light of the independence of the operational definitions of motivation and anxiety, the question arises: why the functional equivalence of these two drives? The conceptions concerning the functional properties of motivation as conceived in this study must be considered inadequate. The evidence suggests that a drive variable that is defined independently of anxiety can also have both positive and negative functional characteristics on the same task. This would be expected, if Duffy's (1957) concept of activation were applicable to different drive variables.

SUMMARY

Seventy-two male, sophomore, college students with middle range scores on the ACE were selected on the basis of operationally defined criteria as being either high or low in Anxiety *and* either high or low on an independent Motivation variable. A series of five perceptual recognition tests were administered individually to all subjects. Half of the subjects in each group were given the tests under conditions designed to be stressful, while half the subjects in each group were given the tests under Nonstress conditions.

In the Stress situation, the HA and the HM groups performed significantly less well than the LA and the LM groups, respectively. In the Nonstress situation, however, the LA and the LM groups performed significantly less well than the HA and the HM groups, respectively. The weight of evidence from this and many other studies bearing on this problem points to the need to include the task and situational variables as a crucial factor in formulating a comprehensive theory of motivation.

REFERENCES

DUFFY, ELIZABETH. The psychological significance of the concept of "arousal" or "activation." *Psychol. Rev.*, 1957, **64**, 265–275.

GORDON, E. M., & SARASON, S. B. The relationship between test anxiety and other anxieties. *J. Pers.*, 1955, **23**, 317–323.

JACKSON, D. N. A short form of Witkin's Embedded Figures Test. *J. abnorm. soc. Psychol.*, 1956, **53**, 254–255.

MANDLER, G., & COWEN, JUDITH E. Test Anxiety Questionnaires. *J. consult. Psychol.*, 1958, **22**, 228–229.

MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.

MATARAZZO, J. D., & PHILLIPS, J. S. Digit symbol performance as a function of increasing levels of anxiety. *J. consult. Psychol.*, 1955, **19**, 131–134.

MATARAZZO, J. D., ULETT, G. A., & SASLOW, G. Human maze performance as a function of increasing levels of anxiety. *J. gen. Psychol.*, 1955, **53**, 79–95.

MATARAZZO, R. G., & MATARAZZO, J. D. Anxiety level and pursuitmetor performance. *J. consult. Psychol.*, 1956, **20**, 70.

ROCKETT, F. C. Form recognition as a function of test-anxiety, color figure, and color-pattern. Unpublished doctoral dissertation, University of Texas, 1955.

RUEBUSH, B. K. Interfering and facilitating effects of test anxiety. *J. abnorm. soc. Psychol.*, 1960, **60**, 205–212.

SARASON, I. G. Effect of anxiety, motivational instructions, and failure on serial learning. *J. exp. Psychol.*, 1956, **51**, 253–260.

SARASON, I. G. Effect of anxiety and two kinds of motivating instructions on verbal learning. *J. abnorm. soc. Psychol.*, 1957, **54**, 166–171. (a)

SARASON, I. G. Test anxiety, general anxiety, and intellectual performance. *J. consult. Psychol.*, 1957, **21**, 485–490. (b)

SARASON, I. G. Intellectual and personality correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1959, **59**, 272–275.

SARASON, S. B., & MANDLER, G. Some correlates of text anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 810–817.

THURSTONE, L. L. *A factorial study of perception.* Chicago: Univer. Chicago Press, 1944.

WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, PEARL B., & WAPNER, S. *Personality through perception.* New York: Harper, 1954.

(Received February 14, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 64, No. 3, 222–228

SIGNIFICANT FACTORS IN HYPNOTIC BEHAVIOR[1]

THEODORE XENOPHON BARBER AND LOUIS BENJAMIN GLASS[2]

State Hospital, Medfield, Massachusetts

Wells' (1924) informal studies on "waking hypnosis" suggest the following hypotheses:

1. Direct suggestions, such as "Your body is immovable," or "You cannot say your name," administered *without* a preliminary trance induction procedure, are sufficient to produce body immobility, verbal inhibition, and other behaviors historically associated with the word "hypnosis" in a "considerable proportion" of subjects.

2. Such "hypnotic behaviors" can be as effectively elicited by direct suggestions given alone as by suggestions given after a formal trance induction.

The results of the study of Weitzenhoffer and Sjoberg (1961) tend to support the first hypothesis but not the second. These investigators found that suggestions given after a formal induction procedure were more effective than the same suggestions given alone in producing arm rigidity, eye catalepsy, visual-auditory hallucination, negative hallucination, and other behaviors termed "hypnotic." There is additional confirmation of the first hypothesis in the recent studies of Barber (1959), Barber and Deeley (1961), and Barber (1961).

The purpose of the present investigation was twofold: to provide a further test of Wells' (1924) hypotheses (Studies I and II) and to explore the possible biographical and attitudinal factors which may be associated with responsiveness to direct suggestions (Study III).

STUDY I. PARAMETERS OF "SUGGESTIBILITY" IN A STUDENT POPULATION

Method

Subjects. All subjects resided in college dormitories and were obtained by a room-to-room approach in which they were asked to participate in an experiment on imagination and concentration (hypnosis was not mentioned). Of 497 students approached, 462 (93%) agreed to participate. The 462 participants consisted of 236 women at Regis College, and 123 women and 103 men at Boston University. The age range was 17–28 with a median of 19.

Procedure. All subjects were tested by the same experimenter.[3] Preliminary instructions were:

> These are all tests of imagination. The better you can imagine and the harder you try, the more you'll respond. Try as hard as you can to concentrate, and to imagine the things I tell you.

As an introduction to the tasks, the Chevreul Pendulum Test, as described by Weitzenhoffer (1953, p. 12), was administered. Each subject was then seated in a comfortable chair and given the following eight tests (direct suggestions) in the same way, as memorized by the experimenter; the first three tests were given with the subject's eyes open, the remaining five with eyes closed.

1. Arm Lowering. Starting with the subject's right arm extended and horizontal, suggestions are given for 30 seconds that the arm is becoming heavy and is moving down. Scoring criterion: 1 point for response of 4 inches or more.

2. Arm Levitation. Starting with the subject's left arm extended and horizontal, suggestions are given for 30 seconds that the arm is weightless and is moving up. Scoring criterion: 1 point for response of 4 inches or more.

3. Hand Lock. Subject is instructed to clasp his hands together tightly with fingers intertwined and place them in his lap. Suggestions are given for 45 seconds that the hands are like steel, they are welded together, they cannot be taken apart. Scoring criteria: ½ point for incomplete separation of hands after 5-seconds effort; 1 point for incomplete separation after 15-seconds effort.

4. Thirst "Hallucination." Suggestions of extreme thirst are given for 45 seconds. Scoring criteria: ½ point if the subject shows noticeable swallowing, moistening of lips, or marked mouth movements; additional ½ point if the subject indicates (during the post-experimental interview) that he became thirsty during this test.

5. Verbal Inhibition. Suggestions are given for 45 seconds that the subject's throat and jaw muscles are rigid and he cannot say his name. Scoring criteria: ½ point if the subject does not say his name after 5-seconds effort; 1 point if he does not say his name after 15-seconds effort.

[1] This investigation was supported by a research grant (MY3253) from the National Institute of Mental Health, of the National Institutes of Health, United States Public Health Service.

The authors wish to thank Joan Burke for administering the tests in Study I and Milton S. Palanker for obtaining the male subjects.

[2] Now at the University of Michigan.

[3] Joan Burke.

TABLE 1

SUMMARY OF ANALYSIS OF VARIANCE OF SCORES ON EIGHT DIRECT SUGGESTIONS ATTAINED BY THREE GROUPS OF MALE AND FEMALE COLLEGE STUDENTS

Source of variation	*SS*	*df*	*MS*	*F*
Between groups	10.55	2	5.27	1.09[a]
Within groups	2208.81	459	4.81	
Total	2219.36	461		

[a] Not significant.

6. Body Immobility. Suggestions are given for 45 seconds that the subject's body is heavy and rigid and he cannot stand up. Scoring criteria: ½ point if the subject is not completely standing after 5-seconds effort; 1 point if not completely standing after 15-seconds effort.

7. "Posthypnotic" Response. Subject is told:

> When this experiment is over I'll click like this [the experimenter presents auditory stimulus] and you'll cough automatically. At the moment I click you'll cough.

Scoring criterion: 1 point if the subject coughs in response to auditory stimulus.

8. Selective Amnesia. Subject is told that when the experiment is over he will remember all the tests except the one where he was told his arm was moving up (Test 2), and that he will remember this test only when the experimenter says, "Now you can remember." Scoring criterion: 1 point if the subject does not refer to amnesic task but recalls all other tests through Test 6 and then recalls Test 2 in response to cue words.[4]

Following the administration of the tests, the subject was informed that the experiment was over and told to open his eyes. To score Test 8 the subject was asked to recite the tests; to score Test 7 the auditory stimulus was presented at an appropriate interval. The subject was then requested to describe his experiences during the testing; following this, he was asked not to discuss the experiment with others.

All subjects tested during the latter part of the summer of 1960 (60 women at Boston University) were given a retest (by the same experimenter) after an interval of two days to one week; the other 402 subjects (tested during the winter, spring, early summer, and fall of the same year) were given the tests once.

Results and Discussion

The mean scores on the first testing for the three groups (women at Regis College, women at Boston University, and men at Boston University) are 3.45, 3.10, and 3.24. Since an analysis of variance, as summarized in Table 1, indicates that the scores attained by the three groups do not differ significantly, the scores for the entire sample are combined in Table 2. As this table indicates, 102 subjects (22%) showed a "high" level of response to the tests—scores of 5.5 or above on an eight-point scale. The post-experimental reports of these subjects were comparable to the reports one might expect from persons who have been formally "hypnotized"; e.g., "I felt I was dying from thirst," "I was amazed when I couldn't speak my name," "I just couldn't get up from the chair." As Table 2 also indicates an additional 231 subjects

TABLE 2

SCORES OF 462 COLLEGE STUDENTS ON EIGHT DIRECT SUGGESTIONS

Level of suggestibility	Raw score	Number of subjects		Percentage of subjects		Centile equivalent	*T* score
High	8	12	102	2.6	22	99	72
	7.5	8		1.7		96	68
	7	16		3.5		94	65
	6.5	17		3.7		90	63
	6	26		5.6		86	60
	5.5	23		5.0		80	58
Medium	5	24	231	5.2	50	75	56
	4.5	28		6.1		70	55
	4	39		8.4		62	53
	3.5	27		5.8		55	51
	3	44		9.5		48	49
	2.5	31		6.7		40	47
	2	38		8.2		32	45
Low	1.5	25	129	5.4	28	25	43
	1	35		7.6		19	41
	0.5	25		5.4		12	38
	0	44		9.5		5	33
Total		462		100			

[4] Subjects at Boston University were given Tests 7 and 8 as described above (Form B). The following alternate form (Form A) of Tests 7 and 8 had been administered to subjects previously tested at Regis College. Test 7 ("Posthypnotic" Response): subject was told he would blink each time the auditory stimulus was presented (the stimulus was presented three times). Test 8 (Selective Amnesia): subject was told he would not remember Test 4 (Thirst "Hallucination"). Since the blinking response (Test 7, Form A) was at times difficult to score and since selective amnesia for the thirst "hallucination" (Test 8, Form A) at times interfered with scoring Test 4, Form B of Tests 7 and 8, which could be scored without difficulty and without interference with earlier tests, was administered to subjects subsequently tested at Boston University.

(50%) passed from two to five tests, and the remaining 129 subjects (28%) passed less than two tests.

The number of subjects passing (scoring 1 point) on each test was: Arm Lowering—270 (58%), Arm Levitation—226 (49%), Hand Lock—96 (21%), Thirst "Hallucination"—277 (60%), Verbal Inhibition—69 (15%), Body Immobility—45 (10%), "Posthypnotic" Response—208 (45%), Selective Amnesia—63 (14%). If 45% represents a "considerable proportion" of subjects, the first hypothesis—that behaviors associated with the word "hypnosis" can be induced in a "considerable proportion" of subjects by direct suggestions without a preliminary trance induction procedure—holds, in the present sample, for arm lowering, arm levitation, thirst "hallucination," and "posthypnotic" response. Although suggestions of verbal inhibition, body immobility, and selective amnesia are rated on some scales of "hypnotic depth" (cf. LeCron & Bordeaux, 1949, pp. 65–66) as "difficult" suggestions which indicate a "medium" or "deep trance," a small minority of subjects (10–15%) in the present sample passed these tests without a formal trance induction.

The 60 subjects given a retest showed lower scores in the second session: the mean score on the first testing (3.13) is significantly higher than the retest score (2.58) ($t = 3.35$, $p < .01$). A product-moment correlation coefficient of 0.88 ($p < .001$) was obtained between scores in the two sessions.

The results of this investigation are in line with Weitzenhoffer and Sjoberg's (1961) study which found that "some individuals can produce hypnotic-like behavior in the absence of any formal induction of hypnosis." Previous studies from our laboratory suggest a similar conclusion. Barber and Deeley (1961) found that some unhypnotized persons who are instructed to "try not to see the color red and the color green" give responses to the Ishihara Plates similar to those given by "hypnotic color-blind" subjects. Oswald (1957) and Barber (1959) have presented data indicating that a performance considered characteristic of the "deeply hypnotized" person—"hallucinating" colors on a gray card, and subsequently reporting appropriate complementary colored afterimages (Erickson & Erickson, 1938; Rosenthal & Mele, 1952)—can be elicited from some persons (without a preliminary trance induction) by instructions to imagine the specified colors. Additional investigations, recently reviewed elsewhere (Barber, 1961) suggest that a trance induction procedure is not necessary to elicit other behaviors associated with the term "deep hypnosis" such as hypnotic deafness, hypnotic blindness, and negative hallucinations; it appears likely that similar performances can be induced in some normal persons by the following brief instructions:

> Try to remain inattentive to the auditory (or visual) stimuli.

STUDY II. THE RELATIVE EFFECTIVENESS OF DIRECT SUGGESTIONS AND A CONVENTIONAL TRANCE INDUCTION IN ELICITING "HYPNOTIC BEHAVIORS"

The present study was designed to test the second hypothesis: Direct suggestions given alone are as effective as the same suggestions given after a formal trance induction in eliciting behaviors associated with the word "hypnosis."

Method

Subjects. Thirty subjects (20 men and 10 women), who had not been previously hypnotized, volunteered to participate. Fourteen were male college students and 16 were adults residing at a summer camp. The age range was 18–41 with a median of 24.

Procedure. Subjects were informed in advance that two experimental sessions were planned, the first session involving only tests of imagination given under normal conditions, and the second session involving hypnosis.

Instructions for the first session (control) were:

> Today you will be given tests of imagination under normal conditions. You will not be hypnotized. The better you can imagine and the harder you try, the more you'll respond; try as hard as you can to concentrate, and to imagine the things I tell you.

The seated subject was instructed to close his eyes and the eight tests (direct suggestions) were administered as described in Study I above.[5]

[5] In this experiment, Form A of Test 7 and 8, as described in Footnote 4, was used, and the Chevreul Pendulum was not employed as a preliminary test.

The second session (formal trance induction) was held after a lapse of not less than two days or more than one week. Preliminary instructions were:

> In the previous session, I tested your ability to imagine under normal conditions; today you're going to be deeply hypnotized and I think you'll find it an interesting and enjoyable experience.

The seated subject was instructed to fixate on an electric metronome which ticks in synchrony with a blinking light, and a 20-minute trance induction, patterned after the procedures outlined by Friedlander and Sarbin (1938), Weitzenhoffer and Hilgard (1959), and Marcuse (1959) was administered by an experimenter (LBG) with prior training as a hypnotist. Immediately after the trance induction procedure, the eight tests were administered (as if they were part of the induction procedure) and scored in the same way as in the preceding (control) session.

Results and Discussion

The test-retest results from Study I indicate that if the eight tests are administered twice (without a preliminary trance induction) the mean score tends to drop (by about 0.5 points) in the second session; in the present study, with a formal trance induction procedure preceding the tests in the second session, the average score not only did not drop the second time but was higher (by 0.5 points) than the average score in the first session. A comparison of the scores under the control and experimental conditions using Wilcoxon's matched-pairs signed-ranks test showed that the latter were significantly higher than the former ($T = 33.5$, $N = 19$, $p < .02$).[6] The second hypothesis thus appears untenable: suggestions given after a conventional trance induction procedure are more effective than the same suggestions given alone in eliciting behaviors associated with the word "hypnosis."

Although the trance induction procedure was associated with a higher mean score, 4 subjects (13%) attained higher scores under the control condition (first session), 11 (37%) attained the same score under the two conditions, and only 15 subjects (50%) showed higher scores after the trance induction (second session). Of the 15 subjects attaining higher scores under the experimental condition, 7 gained no more than 1 point and 8 (or 27%) showed larger gains (1.5–3 points). A Spearman rank-correlation coefficient of 0.85 ($p < .001$) was obtained between scores in the two sessions.

Weitzenhoffer and Sjoberg (1961) previously reported similar findings. In their experiment, 17 suggestions of graded difficulty were administered to 30 college students in the "waking" condition; when a subject had failed three successive suggestions, a 15-minute trance induction procedure (consisting of eye-fixation and suggestions of relaxation, drowsiness, and sleep) was administered. After the trance induction, the tests were resumed, beginning two items below the first item failed and continuing until the scale was completed or the subject had again failed three successive items. The scores on the 17-point "suggestibility" scale were significantly higher following the trance induction. However, 17 subjects (57%) did not gain in score after the induction and, of the 13 subjects (43%) showing higher scores, only 5 (17%) showed marked gains (5 or more points on the 17-point scale). In brief, the results of the present experiment (Study II) and the results of the experiment previously reported by Weitzenhoffer and Sjoberg are in agreement as follows: (*a*) a conventional trance induction procedure is associated with a statistically significant overall enhancement of "suggestibility" in a group of subjects, (*b*) this overall gain is due to higher scores attained by approximately half of the subjects, and (*c*) only a small proportion of subjects (about one out of five) show relatively large gain.

The trance induction procedure employed in the present experiment (Study II) included the following components: (*a*) The situation was explicitly defined as hypnosis and implicitly defined as different, as important, as a situation in which maximal responsiveness and unusual behavior was expected. The subject (*b*) fixated on a blinking light while listening to the repetitive tick of the metronome, (*c*) remained relatively immobile with eyes closed for 20 minutes, and (*d*) listened to the hypnotist's statements to the effect that he was becoming relaxed, drowsy, and

[6] Application of the Wilcoxon signed-ranks test to the eight tests individually showed no significant difference between scores under experimental and control conditions on any one test.

sleepy. Could heightened response to suggestions have been obtained if the preliminary procedure was limited only to the first of these components—i.e., if the situation was defined either as hypnosis or as a very important test situation in which behavior normally considered odd or inappropriate was desired and expected—without employing Components *b*, *c*, and *d*? An affirmative answer to this question is suggested by a follow-up experiment (Glass & Barber, 1961) which appears to indicate that, with at least some subjects, a placebo administered by a physician as a "hypnosis producing" drug is as effective as a formal 20-minute trance induction procedure in enhancing response to the eight tests (direct suggestions) described in Study I above. Further experiments are in order to delineate which of the factors subsumed under the term "trance induction procedure" are necessary, which helpful but unnecessary, and which superfluous to producing enhanced response to suggestions.

STUDY III. AN EXPLORATION OF PERSONALITY CHARACTERISTICS RELATED TO "SUGGESTIBILITY"

Our recent experimental studies (Barber, 1956, 1957a, 1957b, 1958, 1960) and a series of earlier investigations reviewed by Weitzenhoffer (1953) suggest a possible relationship between the characteristics of the subject, as a person, and responsiveness to suggestions of the type commonly employed in hypnosis experiments. This indicated relationship was explored in the present study.

In a prior investigation with 70 subjects (Barber, 1960), 61 items from three questionnaires (the Guilford-Zimmerman Temperament Survey, the Webster-Sanford-Freedman (1955) version of the F Scale, and a 14-item questionnaire especially constructed for the study) were found to differentiate "highly suggestible" from "unsuggestible" subjects. A questionnaire consisting of these 61 items was administered to the three groups participating in Study I (women at Regis College, women at Boston University, and men at Boston University), and to the single group participating in Study II. Analysis of the results for the 106 "highly suggestible" and the 140 "unsuggestible" subjects [7] revealed four items which differentiated the latter from the former *in all four groups* and which were significant at the .02 level by the chi square test. To determine if these four items would "hold-up" in a previously untested group, an additional group (20 women college students) was given the questionnaire and rated on the eight tests (direct suggestions) as described in Study I (by the same experimenter who had carried out Study I). The three "highly suggestible" and the six "unsuggestible" subjects in the new group answered each of the four items in the predicted direction. The number of positive items, the content of the items, and the fact that each item held up over five groups tested by two experimenters is sufficient to exclude chance findings (Brozek & Tiede, 1952). The differentiating items are as follows ("highly suggestible" subjects tend to answer Yes; "unsuggestible" subjects tend to answer No):

1. You like to sell things (that is, act as a salesman) ($\chi^2 = 7.08$, $p < .01$).
2. You like to read true stories about love and romance ($\chi^2 = 11.07$, $p < .001$).
3. You find daydreaming very enjoyable ($\chi^2 = 6.05$, $p < .02$).
4. When you were a child of about five or six, did you have imaginary playmates who were rather vivid and almost real? ($\chi^2 = 20.86$, $p < .001$.)

In a follow-up interview, the "highly suggestible" subjects (in Study I) who had given a Yes answer to Item 1 were asked, "*Why* do you like to sell things (that is, act as a salesman)?"; the answers obtained (e.g., "Because I like people," "I like to meet new people") suggest that Item 1 refers to readiness to form new interpersonal relationships. Items 2–4 appear to refer to a pattern of behavior which can be provisionally categorized as proneness to imaginative activities, daydreaming, and fantasy.

Although the four items "held-up" over the five groups tested in these experiments, the following should be noted:

[7] Subjects attaining scores of 5.5 or above on the eight tests (direct suggestions) are categorized as "highly suggestible"; subjects scoring below 2 are categorized as "unsuggestible."

One of the "highly suggestible" subjects ($N = 107$) and one of the "unsuggestible" subjects ($N = 141$) did not complete the questionnaire.

1. Only two experimenters participated and, with the exception of 16 subjects tested in Study II, all subjects were college students; it remains to be determined if the items apply to nonstudent subjects and to subjects tested by different experimenters.

2. The items differentiate only the "highly suggestible" from the "unsuggestible" subjects and do not apply to the large group rated medium on "suggestibility."

3. Since only 4 of the original 61 items "held" over the five groups tested by two experimenters, it appears possible that "suggestibility" is not in one-to-one correspondence with fixed "personality characteristics" but is multidetermined—by situational factors, by the characteristics of the experimenter, by an interaction between the subject's characteristics and situational factors, and by an interaction between the subject's characteristics and the experimenter's characteristics.

SUMMARY

The primary purpose of this investigation was to test two hypotheses suggested by hypnosis": (*a*) Direct suggestions (administered without a preliminary trance induction procedure) are sufficient to elicit behaviors associated with the word "hypnosis" from a "considerable proportion" of subjects, and (*b*) behaviors associated with the word "hypnosis" can be as effectively elicited by direct suggestions given alone as by suggestions given after a formal trance induction procedure.

In an investigation (Study I) concerned with the first hypothesis, three groups of male and female college students ($N = 462$) were given eight direct suggestions designed to elicit the following items of behavior: arm lowering, arm levitation, inability to unclasp hands, thirst "hallucination," inability to say name, body immobility, selective amnesia, and "posthypnotic" response. Although a formal trance induction procedure was not employed, 102 subjects (22%) carried out more than five of the eight suggested behaviors. The experimental behaviors and post-experimental reports of these subjects were comparable to what one might expect had they been formally "hypnotized." An additional 231 subjects (50%) responded to between two to five suggestions, and the remaining 129 subjects (28%) responded to less than two suggestions. If 45% represents a "considerable proportion" of subjects, the first hypothesis was confirmed, in this sample, for the following behaviors: arm lowering, arm levitation, thirst "hallucination," and "posthypnotic" response.

In an experiment (Study II) designed to test the second hypothesis, 30 subjects were given the eight direct suggestions twice: first, without a preliminary trance induction procedure (control condition) and, second, following a formal trance induction (experimental condition). The second hypothesis was not confirmed: in harmony with a previous study reported by Weitzenhoffer and Sjoberg (1961), it was found that a trance induction procedure had a significant overall effect in enhancing response to suggestions. Fifteen subjects (50%) attained higher scores under the experimental condition, 11 (37%) attained the same score under the two conditions, and 4 (13%) attained higher scores under the control condition.

A concomitant study (Study III) was carried out to explore an indicated relationship between personality characteristics and responsiveness to suggestions of the type commonly employed in hypnotic experiments. A questionnaire constructed in an earlier investigation (Barber, 1960) was administered to the three groups rated on "suggestibility" in Study I, to the group rated in Study II, and to an additional group of 20 subjects rated in a follow-up study. The results suggest that in this sample of American college students the "highly suggestible" subjects, as compared with the "unsuggestible," are more willing to form new interpersonal relationships, and are more prone to imaginative activities, daydreaming, and fantasy.

REFERENCES

BARBER, T. X. A note on "hypnotizability" and personality traits. *J. clin. exp. Hypnosis*, 1956, **4**, 109–114.

BARBER, T. X. Experiments in hypnosis. *Scient. American*, 1957, **196**(4), 54–61. (a)

BARBER, T. X. Hypnosis as perceptual-cognitive restructuring: I. Analysis of concepts. *J. clin. exp. Hypnosis*, 1957, **5**, 147–166. (b)

BARBER, T. X. The concept of "hypnosis." *J. Psychol.*, 1958, **45**, 115–131.

BARBER, T. X. The after-images of "imagined" and "hallucinated" colors. *J. abnorm. soc. Psychol.*, 1959, **59**, 136–139.

BARBER, T. X. The necessary and sufficient conditions for hypnotic behavior. *Amer. J. clin. Hypnosis*, 1960, **3**, 31–42.

BARBER, T. X. Physiological effects of "hypnosis." *Psychol. Bull.*, 1961, **58**, 390–419.

BARBER, T. X., & DEELEY, D. C. Experimental evidence for a theory of hypnotic behavior: I. "Hypnotic color-blindness" without "hypnosis." *Int. J. clin. exp. Hypnosis*, 1961, **9**, 79–86.

BROZEK, J., & TIEDE, K. Reliable and questionable significance in a series of statistical tests. *Psychol. Bull.*, 1952, **49**, 339–341.

ERICKSON, M. H., & ERICKSON, ELIZABETH, M. The hypnotic induction of hallucinatory color vision followed by pseudo-images. *J. exp. Psychol.*, 1938, **22**, 581–588.

FRIEDLANDER, J. W., & SARBIN, T. R. The depth of hypnosis. *J. abnorm. soc. Psychol.*, 1938, **33**, 281–294.

GLASS, L. B., & BARBER, T. X. Hypnotic behavior, the definition of the situation, and the placebo-effect. *J. nerv. ment. Dis.*, 1961, **132**, 539–541.

LECRON, L. M., & BORDEAUX, J. *Hypnotism today*. New York: Grune & Stratton, 1949.

MARCUSE, F. L. *Hypnosis—fact and fiction*. Baltimore: Penguin, 1959.

OSWALD, I. After-images from retina and brain. *Quart. J. exp. Psychol.*, 1957, **9**, 88–100.

ROSENTHAL, B. G., & MELE, H. The validity of hypnotically induced color hallucinations. *J. abnorm. soc. Psychol.*, 1952, **47**, 700–704.

WEBSTER, H., SANFORD, N., & FREEDMAN, M. A. new instrument for studying authoritarianism in personality. *J. Psychol.*, 1955, **40**, 73–84.

WEITZENHOFFER, A. M. *Hypnotism—an objective study in suggestibility*. New York: Wiley, 1953.

WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford hypnotic susceptibility scale*. Palo Alto, Calif.: Consulting Psychologists Press, 1959.

WEITZENHOFFER, A. M., & SJOBERG, B. M., JR. Suggestibility with and without "induction of hypnosis." *J. nerv. ment. Dis.*, 1961, **132**, 204–220.

WELLS, W. R. Experiments in waking hypnosis for instructional purposes. *J. abnorm. soc. Psychol.*, 1924, **18**, 389–404.

(Received February 23, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 64, No. 3, 229–233

CRITIQUE AND NOTES

THE RELATIONSHIP BETWEEN RESPONSE STYLE AND PERSONALITY VARIABLES:

I. THE MEASUREMENT OF RESPONSE ACQUIESCENCE [1]

RICHARD K. McGEE [2]

Moccasin Bend Psychiatric Hospital, Chattanooga, Tennessee

A recent approach to the response set component of test scores is to treat it not as error variance, but as a personal stylistic variable of interpretive significance in personality assessment (Jackson & Messick, 1958). Much attention has been focused on *response acquiescence*, the tendency to agree with test items, regardless of their content. Couch and Keniston (1960) presented the most recent data to confirm the general proposition that the "agreeing response tendency is based on a central personality syndrome" (p. 173). Typically, this conclusion is reached after the demonstration of a relationship between a single measure of acquiescence and some measure of conformity, authoritarianism, rigidity, dependency, anal resentment, or low ego strength, to name only a few. McGee (1962) has reviewed these studies elsewhere, and discussed some methodological revisions for the design of future investigations.

It may be observed that each of the major investigations of the personality correlates of response acquiescence has used a different instrument for measuring the response style. It would be valuable to know how much variance, if any, these measures have in common. To hypothesize that response style has its roots in the responder's personality implies an assumption that the response tendency is a generalized behavior which may be expected to appear in a variety of "agree-disagree" situations. Unless this assumption can be verified, it would be difficult to defend the position that a personal mode of responding springs from a central personality syndrome.

There is general agreement in the literature that the verbal content of items is relatively unimportant; that acquiescent individuals respond "Yes," or "Agree" irrespective of item content. Jackson and Messick (1958) reviewed the evidence, and concluded that "the *major common factors in personality inventories of the true-false or agree-disagree type . . . are interpretable primarily in terms of style rather than specific item content*" (p. 247). What is important is the degree of structure, or ambiguity, of the stimuli. Cronbach (1946) stated that response sets are most likely to appear when the stimuli are ambiguous. Berg and Rapaport (1954) confirmed this prediction by showing that an "unstructured questionnaire," in which subjects responded to imaginary items, yielded a consistent tendency for acquiescent response options to be selected.

The foregoing considerations suggest the major hypothesis which it is the purpose of this study to test: Response acquiescence is a general test taking behavior which appears in any situation characterized by the forced choice of "agree-disagree" response options. This hypothesis may be examined by observing the appearance of response acquiescence when specific variations are made in the stimulus conditions, i.e., the item content. For example, if stylistic determinants, rather than content determinants account primarily for the variance of test scores, high correlations should be found between two measures of acquiescence, even though they differ markedly from one another in the meaningfulness of their verbal content. Similarly, if response styles appear most readily in ambiguous situations, high correlations should be found between measures of acquiescence which contain highly ambiguous stimuli.

METHOD

Six tests of acquiescence were administered to 218 undergraduate students in general psychology classes. The sample included 144 males and 74 females, ranging in age from 17 to 24 years. Five of the tests were taken in large group sessions approximating 100 subjects each, and lasting about 1 hour. The sixth test was presented to groups of about 40 subjects each in 20-minute sessions.

For the larger sessions, the five tests were mimeographed and assembled into booklets. Order of presentation was varied by using 10 random arrange-

[1] The author is indebted to Samuel S. Komorita and Jum C. Nunnally for their aid in planning this study and preparing this report of the results.

[2] This investigation was completed while the author was at Vanderbilt University.

ments, and subjects were told to complete the forms as they appeared in their booklet. The sixth test consisted of stimulus pictures which were projected by an opaque projector onto a screen at the front of the classroom. This form was taken by approximately half of the subjects prior to the group of five tests, and afterwards by the remaining half.

The six tests used in this study were selected because they represent the major techniques used thus far in the research on acquiescence. Three of them contain items loaded with meaningful verbal content, whereas the other three are essentially content free. The content tests used were the Social Acquiescence Scale (SAS), the California F Scale, and the Agreement Response Scale (ARS). Bass (1956) constructed the 56-item SAS out of proverbs, aphorisms, and other famous sayings. The F Scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950) is well known for its use as an attempt to measure both acquiescence and authoritarianism. The 30-item Form 40–45 was used here. The ARS was published by Couch and Keniston (1960) as the best short form of their 360-item test of the agreeing response tendency. Each of the 15 items correlates .40 or better with the general agreement factor in their data.

The three noncontent tests used have been devised for the purpose of measuring "pure" acquiescence by forcing either agreement or disagreement responses to unstructured, or meaningless stimuli.[3] The subjects were told they were participating in three different experiments: A study of the recognition value of German words, a study of the effectiveness of Extra-Sensory Perception, and a study of the amount of information communicated by abstract art. The German Language Recognition Form (Nunnally & Husek, 1958) contains 24 sentences consisting of two German nouns coupled by English verbs and modifiers. The ESP scale consisted of 25 partial sentences which included only two- or three-word stems, an indication of the length of the sentence, and the final word. It was presented as a model of the classic ESP studies in which the real, complete sentences were "locked up in a safe downstairs." The subjects were to concentrate on the (imaginary) list and attempt to read enough of the statement to indicate their agreement or disagreement with it. The Ambiguous Pictures scale consisted of 20 meaningless designs each drawn with black india ink on pieces of white paper 8 inches square. Beneath each of these designs was printed the title of a popular mystery story. The subjects were told this was a test of the information transmitting properties of abstract art.

For all of the noncontent tests the instructions urged that even though the items appeared silly, or meaningless, the responses would be useful; that the desired experimental effects could be reliably measured if the subjects indicated whether they agreed or disagreed with the statements (German sentences and ESP sentences) or with the choice of title for the pictures; and it was suggested that subjects should obey their first impulse in responding.

Agreement responses were recorded in the same manner for all six of the tests. Mimeographed answer sheets were distributed which contained 60 seven-point rating scales. The subjects checked their degree of agreement with each item according to the following anchors: Completely Disagree, Mostly Disagree, Slightly Disagree, Neutral, Slightly Agree, Mostly Agree, and Completely Agree. Instructions explicitly stated that the Neutral category was to be used only rarely, if at all; that a decision should be forced in one direction or the other. Each of the tests was scored by summing the scale values of the ratings over all the items.

RESULTS

Table 1 shows the matrix of coefficients obtained for the six acquiescence scores and sex. Scatter diagrams indicated that the assumption of linear regression was not violated in the data, thus the product-moment formula was applied for the interscale correlations. The degree of relationship between the tests and sex was computed by the point biserial formula such that a positive correlation indicates that the mean for males is greater than the mean for females.

Corrected split-half (odd-even) reliability coefficients are shown in parentheses along the diagonal. It should be noted that the reliability of the SAS is within the range of values (.81–.92) which Bass (1956) reported for the instrument. The reliability estimate for the F Scale falls slightly below the .81–.89 range reported by Adorno et al. (1950) for the combined Form 40–45. However, variations in the size and other characteristics of the sample must be considered in making such a comparison. The important point is that only two of the six acquiescence scores appear to possess adequate reliability even for research instruments.

An important finding regarding the reliability of noncontent scales appears in connection with the German Language Form. Nunnally and Husek (1958) found substantial correlations (−.69 for example) between these same items and variables such as age and education. Therefore, the scale must have considerably greater reliability when

[3] Copies of the three noncontent test stimuli may be obtained without cost from the author. The ESP scale is similar in format to the "unstructured questionnaire" used by Berg and Rapaport (1954), however the present form of this technique was developed independently by Nunnally. The Ambiguous Pictures were constructed from the designs which form one subtest of Yacorzynski's battery for detecting brain damage, and from random combinations of items in the PRT (Berg, Hunt, & Barnes, 1949). These materials were available through the assistance of K. Edward Renner of Northwestern University, for which the author is grateful.

TABLE 1

INTERCORRELATIONS BETWEEN SIX MEASURES OF ACQUIESCENCE AND SEX
(N for all variables = 218)

Variable	SAS	F Scale	ARS	German words	ESP	Ambiguous pictures	Sex
SAS	(.85)	.56	.51	.27	.27	.01	−.06
F Scale		(.77)	.39	.06	.19	.01	.10
ARS			(.54)	.13	.20	−.03	.12
German words				(.30)	.24	.06	−.20
ESP					(.52)	.18	−.16
Ambiguous pictures						(.57)	−.19
Sex (r_{pbi})							

Note.—r = .14 significant at .05; r = .18 significant at .01.

used with a heterogeneous, noncollege population. However, when these demographic variables are controlled, a major portion of the reliable response set variance is eliminated.

DISCUSSION

How large a correlation coefficient should be before it becomes meaningful is in many ways a value judgment. Many of the values in Table 1 are significantly different from zero at the .01 probability level. Edwards' (1950) distinction between statistical significance and practical significance is relevant here. In the search for sources of common variance a statistically significant correlation is of little practical value if a very small percentage of the variance of one scale is explained by its relationship with the other. The majority of coefficients in Table 1 indicate less than 7% of the variance of one scale is common to them both. Hence, for the most part, these data offer little evidence that a common response tendency was influencing the selection of response options across the six scales.

Other causes of low correlations should be considered in drawing the conclusion stated above. The role of low reliabilities of individual scales in shrinking the correlations is, of course, important. Yet, when these correlations are corrected for attenuation, the interpretations suggested by the data are not changed. Moreover, the finding of low reliability for the response set scores is meaningful in and of itself. Cronbach (1950) has suggested that the most simple and effective demonstration of a response set is derived from assigning a score for the particular response tendency (number of agreement responses, weighted by intensity, in this study) and computing the internal consistency reliability for this score. Following Cronbach's logic, one finds in the present study little evidence that an agreement response style was operating reliably in four of the six measures.

A second possibility is that the method of scoring these scales confounded the agreeing set with the response set to take extreme positions. However, Komorita [4] compared scores derived from seven-point agree-disagree scales with scores derived from dichotomizing the options. His calculations were made on the same data collected for this study, and he found correlations above .90 for every scale so examined.

Finally, since the three noncontent scales were presented as experiments, it is possible that the subjects were led to expect something near a 50–50 split in the "correct" responses to the stimuli, and therefore, restricted their tendency to agree with more than 50% of the items on these measures. To the extent that this is true, it would lower the variance of these scales. However, one would expect this effect equally across the three measures; while it might have lowered the correlations between them and the content scales, it may well have raised the correlations between the noncontent scales. Yet, it is not evident that any tendency was exerting a common influence on the noncontent scales in this study.

The major finding in Table 1 is the apparent importance of meaningful verbal content in the items as a source of common variance. Clearly the SAS, F Scale, and ARS form a factor grouping. None of the other scales correlate with these content measures to the extent present within this grouping. It is also notable that no prominent factor grouping appears among the noncontent tests. This suggests that where subjects are given a chance to respond in an acquiescent manner to different types of ambiguous stimuli, a very small portion of their responses are explainable by a common response tendency. In the correlations across degree of stimulus meaningfulness, the data again fail to support the assumption of a response set variable common to the tests. Cronbach (1950) cites evidence that

[4] Personal communication.

acquiescence scores are correlated across similar situations, i.e., personality inventory items. (Here the average correlation reported is .24.) He also presents evidence that response sets are not consistent over widely varying situations. These data were derived from judgments of physical stimuli as either "pleasant" or "unpleasant." It now appears that when stimuli vary on the content meaningfulness dimension, the tendency to acquiesce likewise fails to appear consistently.

The most parsimonious interpretation of the data in Table 1 is that there is no general trait of response acquiescence which appears reliably in different situations; it is, rather, dependent upon specific groups of items used to measure it. That is, a "pure response set" may be a straw man. This conclusion is clearly suggested by these data, and it receives support from other independent research.

In two recent papers (Jackson & Messick, in press; Messick & Jackson, 1961) emphasis has been placed on variations in the acquiescence-eliciting properties of different item forms. The investigations of these authors (Messick & Jackson, 1961) have lead them to assert that

> the acquiescence-evoking properties of items are not . . . uniform over all scales, but that acquiescence is elicited differentially as a function, perhaps, of specific item content, of the clarity or ambiguity with which the content is stated, and in particular of the perceived desirability of the statement (p. 302).

Therefore, perhaps one should not expect high correlation across scales which vary as grossly in item form as the six measures used in this study. Of course, the corresponding conclusion is that one should not expect to find a general trait of acquiescence either.

Gray and Crisp (1961) have also reasoned that the use of a response set score as a personality trait requires that the score generalize further than a single set of items. They obtained acquiescence responses to a modification of the Perceptual Reaction Test (Berg, Hunt, & Barnes, 1949) and to a completely content free task presented as a measure of subliminal perception. The two sets of scores differed markedly from one another in distribution, in mean acquiescence, and correlated only −.01. Clearly, a common behavioral tendency such as response acquiescence could not explain the responding to these two unstructured, nonverbal measures. They concluded that

> it is still parsimonious and appropriate to regard with suspicion the notion of a generalized free-floating response set utterly divorced from content (p. 11).

The findings of the present study, congruent as they are with other very recent research, have important implications for the new approach to response set scores as traits based on central personality syndromes. It does not appear feasible to continue thinking in terms of a pure response set independent of specific item content. Hence, in attempting to relate acquiescence to personality traits, scores derived from a single instrument are not sufficient to constitute the measure of the response style.

SUMMARY

This study is the first in a series of investigations designed to evaluate the currently popular hypothesis that response styles are based on personality traits, and thereby have utility in personality assessment. Responses of 218 subjects on six different scales of acquiescence, varying in degree of meaningful verbal content, were intercorrelated. The results revealed that only those scales containing similar verbal content in the items were related. Apparently verbal content is quite important, whereas the amount of structure of the items is less important in determining agreement responses, than has been previously supposed.

The data are interpreted as suggesting that there is no general trait of response acquiescence independent of specific instruments used to measure it.

REFERENCES

ADORNO, T. S., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

BASS, B. M. Development and evaluation of a scale for measuring social acquiescence. *J. abnorm. soc. Psychol.*, 1956, **53**, 296–299.

BERG, I. A., HUNT, W. A., & BARNES, E. H. *The Perceptual Reaction Test*. Evanston, Ill.: Berg, 1949.

BERG, I. A., & RAPAPORT, G. M. Response bias in an unstructured questionnaire. *J. Psychol.*, 1954, **38**, 475–481.

COUCH, A., & KENISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *J. abnorm. soc. Psychol.*, 1960, **60**, 151–174.

CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt.*, 1946, **6**, 475–494.

CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt.*, 1950, **10**, 3–31.

EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.

GRAY, C. W., & CRISP, H. E. The credibility of pure response set. Paper read at Southeastern Psychological Association, Gatlinburg, April 1961.

JACKSON, D. N., & MESSICK, S. J. Content and style in personality assessment. *Psychol. Bull.*, 1958, **55**, 243–252.

JACKSON, D. N., & MESSICK, S. J. Acquiescence and desirability as response determinants on the MMPI. *Educ. psychol. Measmt.*, in press.

MCGEE, R. K. Response style as a personality variable: By what criterion? *Psychol. Bull.*, 1962, in press.

MESSICK, S. J., & JACKSON, D. N. Acquiescence and the factorial interpretation of the MMPI. *Psychol. Bull.*, 1961, **58**, 299–304.

NUNNALLY, J. C., & HUSEK, T. R. The phony language examination: An approach to the measurement of response bias. *Educ. psychol. Measmt.*, 1958, **18**, 275–282.

(Received February 9, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 64, No. 3, 234–236

EXPRESSIVE MOVEMENTS, WARMTH, AND VERBAL REINFORCEMENT[1]

MICHAEL M. REECE AND ROBERT N. WHITMAN

Wayne State University

Recent studies have applied the principle of reinforcement in an effort to influence verbalization. Krasner (1958) has reviewed numerous studies in which verbal behavior has been investigated in terms of instrumental learning. A wide variety of stimuli has been found to be effective as secondary reinforcers; e.g., gestures, smiling, nodding, and leaning forward. Verbalization has been shown to have secondary reinforcing properties. Experimenters frequently have used "mm-hmm," "good," and "fine" as a means of reinforcement. These words, and similar ones having the connotation of approval, can be interpreted by the listener as indicating the attitude of the therapist.

Most views of psychotherapy agree on the importance of the therapist's attitude toward the client and the client's interpretation of it. Rogers (1951) has stressed the importance of acceptance, permissiveness, and warmth on the part of the therapist. Fiedler (1953) investigated the influence of the therapist's feelings upon the client and concluded from the evidence he presented that the therapist seems to be "primarily responsible for the character of the therapeutic relationship" (p. 315). It also has been suggested that warmth and acceptance by the therapist can be considered in terms of cues indicative of intense, undivided attention (Krasner, 1955).

In order to be effective, the therapist's attitudes and feelings must be conveyed to the client. They may be communicated verbally and nonverbally. Inferences can be made from the content of the therapist's speech and from the expressive movements which are part of his behavior.

It may be considered that a therapeutic climate of warmth, with the connotation of permissiveness and acceptance, has a reward value for the client, hence a reinforcing effect for learning. The present study was designed to determine whether warmth has a reinforcing influence upon verbal behavior; and particularly, whether warmth combined with verbal reinforcement would produce a greater total reinforcing effect upon verbalization.

[1] Based on an MA thesis by the junior author under the direction of the senior author.

METHOD

The experimental procedure consisted of a 15-minute, "free association" situation. Each subject was asked to say whatever words came to mind and was instructed to say each word individually, not to use phrases or to count. The responses were recorded on tape.

A factorial design was employed with 69 college students who were randomly assigned to four groups designated as warm-reinforced, warm-nonreinforced, cold-reinforced, and cold-nonreinforced.

"Warm" and "cold" were defined in terms of the experimenter's expressive movements, based on a previous study which had related approach and avoidance movements to judgments of warmth and coldness (Reece & Whitman, 1961).

When "warm," the experimenter leaned toward the subject, looked directly at him, smiled, and kept his hands still. When "cold," the experimenter leaned away from the subject, looked around the room rather than at him, did not smile, and drummed his fingers. The expressive movements were maintained continuously for the experimental period. An experienced judge periodically observed the experimenter through a one-way mirror to check the consistency of his movements defining "warm" and "cold."

With the designated groups, plural nouns were reinforced by the experimenter's saying "mm-hmm." Except for the instructions, there was no other verbalization by the experimenter.

Upon completion of the task, each subject completed a questionnaire giving his reactions to the task and to the experimenter, judging him as "warm" or "cold," and indicating the basis for the judgment.

RESULTS

The tapes were replayed and both the frequency of all verbal responses and the frequency of plural noun responses were tabulated. The instructions were followed effectively so that no proper nouns or numbers were given. In addition, the subjects apparently provided their own "set" against repetition of words. Repeated words were scored since repetition rarely occurred.

In order to determine the reliability of this counting procedure, another scorer independently recorded the frequencies of the responses in a sample of 20, randomly selected, 3-minute periods of the tapes. These scores were correlated with those found by the experimenter for the same periods. The Pearson r for all verbal responses

TABLE 1

MEAN TOTAL VERBAL RESPONSES

Group	*N*	*M*	*SD*
Warm-reinforced	16	346.94	114.99
Warm-nonreinforced	15	310.67	117.19
Cold-reinforced	20	298.70	117.01
Cold-nonreinforced	18	298.33	84.93

was .99 and for plural noun responses was .98, indicating very high interscorer consistency.

It is seen from Table 1, which gives the means for all verbal responses, that the warm-reinforced group produced the greatest number of words, and that more verbalization occurred in both warm groups than in the cold groups.

The analysis of variance, shown in Table 2, indicated a significant interaction between verbal reinforcement and expressive movements, and showed that the expressive movements variable also was highly significant.

Marked differences were found with the specifically reinforced plural nouns, as shown in Table 3. The superiority of the reinforced groups was found to be definitely significant by the analysis of variance method (McNemar, 1955) as indicated in Table 4.

Evaluation of the extent to which the subjects' judgments of "warm" and "cold" agreed with the expressive movements defining these conditions yielded a chi square of 20.41 ($n = 3$), significant at less than the .01 level.

When the interview data were examined, it was found that most of the subjects were aware of the expressive movements and utilized them as a basis for their judgments. However, some of the subjects, especially in the cold-reinforced group, appeared to be more responsive to the experimenter's reinforcing verbalization than to the expressive movements.

DISCUSSION

It is evident from the results of this study that the "climate," as defined by expressive movements, was an effective variable. The interaction of warm expressive movements with verbal reinforcement produced the greatest amount of verbalization when all words were considered. Verbal reinforcement alone was not a significant influence upon the total number of spoken words. On the other hand, the expressive movements did prove to have a significant effect.

Since the verbal reinforcement awaited the subjects' responses, the number of reinforcements was not the same for the reinforced groups. Reinforcement therefore would not be expected to produce a uniform result for these groups.

Many studies have shown that verbal reinforcement of a specific class of verbal responses will increase the frequency of those responses. The present findings support and reaffirm this conclusion. The reinforcement of saying "mm-hmm" when plural nouns were voiced was sufficient to produce a significant increase in this response.

Since the warmth or coldness of the experimenter was present throughout a particular condition and was not restricted to plural nouns or specifically associated only with them, it is not surprising that the expressive movements had no significant effect upon this class of response.

It was clearly demonstrated, also, that the expressive movements were not mere manipulations by the experimenter but successfully conveyed the impression of the experimenter's attitude. Leaning toward the subject, smiling, and looking directly at him enabled the subject to judge the experimenter as warm. Conversely, looking away from the subject, leaning away from him, not smiling, and intermittently drum-

TABLE 2

ANALYSIS OF VARIANCE FOR TOTAL VERBAL RESPONSES

Source	*df*	*MS*[a]	*F*
Reinforcement (R)	1	4,694.26	.86
Expressive movements (EM)	1	16,435.62	13.01***
Interaction R × EM	1	5,492.04	4.35*
Within	65	1,263.37	

[a] Corrected for unequal *N* by Snedecor's (1956) method.
* Significant at .05 level.
*** Significant at .001 level.

TABLE 3

PLURAL NOUN RESPONSES

Group	*N*	*M*	*SD*
Warm-reinforced	16	82.44	58.44
Warm-nonreinforced	15	40.00	24.68
Cold-reinforced	20	64.85	40.13
Cold-nonreinforced	18	37.17	30.23

TABLE 4

ANALYSIS OF VARIANCE FOR PLURAL NOUNS

Source	*df*	*MS*[a]	*F*
Reinforcement (R)	1	20,269.14	11.72**
Expressive movements (EM)	1	1,881.24	1.09
Interaction R × EM	1	928.74	.54
Within	65	1,730.15	

[a] Corrected for unequal *N* by Snedecor's (1956) method.
** $p < .01$.

ming the fingers on the table impressed the subject as coldness on the part of the experimenter.

SUMMARY

In order to determine the effects of the experimenter's warmth and coldness upon verbal conditioning in a free association task, 69 subjects were randomly assigned to four experimental groups in a factorial design.

Warmth and coldness were defined in terms of expressive movements of posture, glance, facial expression, and finger activity.

Analysis of the results showed that the total number of words was significantly affected by the expressive movements and verbal reinforcement, but was not significantly influenced by verbal reinforcement alone.

The number of plural nouns was significantly increased by verbal reinforcement but was not significantly affected by the expressive movements or by the interaction of movements and reinforcement.

REFERENCES

FIEDLER, F. E. Quantitative studies on the role of the therapists' feelings toward their patients. In O. H. Mowrer (Ed.), *Psychotherapy: Theory and research*. New York: Ronald, 1953. Pp. 296–325.

KRASNER, L. The use of generalized reinforcers in psychotherapy research. *Psychol. Rep.*, 1955, **1**, 19–25.

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1955.

REECE, M. M., & WHITMAN, R. N. Warmth and expressive movements. *Psychol. Rep.*, 1961, **8**, 76.

ROGERS, C. R. *Client-centered therapy*. New York: Houghton Mifflin, 1951.

SNEDECOR, G. W. *Statistical methods*. Ames: Iowa State Coll. Press, 1956.

(Received January 30, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 64, No. 3, 237–238

OVEREVALUATION OF OWN GROUP'S PRODUCT IN INTERGROUP COMPETITION [1]

ROBERT R. BLAKE AND JANE SRYGLEY MOUTON

University of Texas

Any social system which embraces values such as excellence, efficiency, or adequacy as standards for performance develops normative criteria by which the achievements of individual, groups, or organizations can be evaluated. Depending on results from applying such judgments, the evaluated unit, whether individual, group, or organization, may be accepted or rewarded, rejected or punished, either completely or in varying degrees. The judgments leading to acceptance or rejection may be self-imposed or made by others. Grant that, under such conditions, a fundamental human motive is to increase the probability of acceptance or to avoid the risk of rejection. Then, acceptance generates security and reduces anxiety; rejection increases anxiety and reduces security.

When absolute normative criteria of excellence, efficiency, or adequacy are unavailable, as they frequently are, judgments only can be made in relative terms. With criteria which are relative rather than absolute and with motivation in the direction of increasing security through being excellent, efficient, or adequate then distortions in perception and judgment are possible. Such distortions, it is predicted, are in the direction of evaluating one's own performance above that of a comparison individual or individuals. When the object of evaluation is a *group* product and one values his group membership, the prediction is that distortions of judgments occur in the same direction as above; that is, members evaluate their own group product as being "better" than the product of a comparison group or groups.

The present paper describes an experiment designed to evaluate the accuracy of this prediction under conditions where group products are being compared. The specific hypothesis tested is that group members elevate the evaluation of their own solution to a problem above their evaluation of the product from a comparison group.

PROCEDURE

Setting and Subjects

The experiment was conducted in 17 different laboratory training programs, representing 48 groups and totaling 410 subjects. Competition took place between eight pairs, four trios, and five quartets of groups, with groups composed of 8–10 members each.

Participants in 32 of the 48 groups were adults who ranged in age from 25 to 55 and who were engaged in some aspect of administrative work in connection with industrial, medical, and research organizations. The remaining 16 groups were composed of junior and senior college students, both male and female, enrolled in social psychology. The sequence of activities was the same for adult and college subjects. All groups were matched to equalize membership in terms of educational and occupational level for adult groups and in class level and sex for college groups.

Sequence of Interactions

The arrangements for creating intergroup competition were:

Development of autonomous groups. Ten to 12 hours were devoted to autonomous ingroup activity in which individuals met together to study decision making in groups through developing, and then evaluating, their own ingroup structures.

Creating intergroup competition. After the initial period of ingroup development, an experimental arrangement was created such that groups found themselves in competition with each other. Each group had 3 hours to formulate its own approach to the solution of an assigned problem. The problem, to be solved in the form of a two-page, double-spaced, typewritten memorandum, concerned some issue within an organization or intergroup relations with which all members of the contending groups were equally familiar. Group members routinely accepted the proposition that their performance on the assigned task provided an indication of their adequacy as a problem solving unit. A winner and a loser were to be determined. The possibility of resolving differences between solutions by compromise was unavailable.

Comparison of group positions. Solutions were then dittoed and distributed to all competing group members for purposes of comparison and clarification of differences. Members judged the merits of the contending solutions in private. In some of the competitions, the rating scale read, "How do you feel about the solutions?" with judgments ranging from 9, "best possible," to 1, "worst possible." In others, the scale was, "Rate the solutions in terms

[1] These studies were supported by Grant M-2477, Behavior of Group Representatives under Pressure, National Institutes of Health.

of overall adequacy," from 9, "completely adequate," to 1, "completely inadequate." Findings were unrelated to the scale used.

RESULTS

Results were analyzed in several different ways. One was a comparison of judgments by individuals of own and other group products. Since private individual judgments were made within the context of group interaction the group has been taken as the unit of analysis and group averages of judgments of own and other's products have been compared also. On a group basis, there were 48 possible comparisons between the average group evaluation of own position relative to the proposal created by another. When there were more than two groups in competition with one another, only the highest evaluation of the contenders' products was used for purposes of analysis. For 46 of the comparisons, the product of one's own group was rated higher than the competitor's and in two they were rated equal. In no instance was the average group rating of own solution rated lower than an opponent's solution. The χ^2 of 40.33 ($df = 1$) is significant beyond the .001 level. On an individual-by-individual basis, there were 410 possible comparisons between judgments of one's own group product and that proposed by a contending group. Own group product was rated higher than the comparison group in 378 of the instances, own product was rated as equal in quality 19 times, and in only 13 instances did a group member rate his own group product as inferior to that created by another group. The χ^2 of 291.99 ($df = 1$) also is significant beyond the .001 level.

The average judgment by the 410 subjects of own group product was 7.5 and of the product of the comparison group, 5.6. Where there were three or four groups judging together the highest judgment of the other group's product was used for comparing the mean ratings. The t for the differences is 23.84 ($p < .001$, $df = 409$). The comparable averages computed on a group basis were 7.6 for own group product, and 5.5 for the product created by the comparison group, with a t of 3.21 ($p < .01$, $df = 47$). The results demonstrate that group members uniformly judge their own solutions as being better in quality than they do the solutions of comparison groups.

DISCUSSION AND CONCLUSIONS

Four hundred and ten persons, participating as members of 48 groups in human relations training programs, engaged in a parallel sequence of ingroup and intergroup activities. Groups were placed in competition in pairs, trios, and quartets of groups, with the outcome determined by the quality of the solutions created by the groups. Quality judgments were made by group members on a private, or personal, basis for the adequacy of their own and of the other group's positions. They resulted in one's own product being rated higher than that created by a competitor's.

The results confirm the prediction that group members evaluate their own group product above the judgments they accord to the proposal from a comparison group.

The overevaluation of one's own group product relative to a comparison group can be interpreted in several ways: (*a*) as due to perceptual distortion stemming from group identification and needs in a situation where personal adequacy via group adequacy serves as a criterion of acceptance or rejection under win-lose conditions; (*b*) as stemming from distortions in reporting evaluations for the instrumental purpose of "winning," or because of greater familiarity with the rationale, reasons, and premises of one's own group's solution.

(Received February 2, 1961)

Journal of Abnormal and Social Psychology
1962, Vol. 64, No. 3, 239-240

TIME ORIENTATION IN MALE AND FEMALE JUVENILE DELINQUENTS [1]

ANTHONY DAVIDS, CATHERINE KIDDER, AND MELVYN REICH

Brown University and Emma Pendleton Bradley Hospital

A recent review by Wallace and Rabin (1960) indicates a long history of psychological interest in temporal experience as well as an increasing amount of research activity in this area of investigation. In this journal, Barndt and Johnson (1955) reported that delinquent boys are more present oriented than are nondelinquent boys. In their study, the delinquent and nondelinquent subjects were well-matched on variables such as age, IQ, school achievement, and socioeconomic status. Utilizing a story completion technique, they found that the delinquent boys produced stories with significantly shorter time spans than did the boys in the control group.

A later investigation, reported in this journal by Davids and Parenti (1958), found no differences between emotionally disturbed and normal 11-year-old boys with respect to time orientation. However, both the disturbed and normal subjects in this study were found to be more present oriented than the normal adolescents in the Barndt and Johnson (1955) investigation, but were not different from the adolescent delinquent boys in this respect. Since the Davids and Parenti study utilized the identical story completion technique and scoring procedures from the Barndt and Johnson study, it was possible to make these direct comparisons between findings obtained from different age subjects who were studied in independent investigations conducted in quite different institutional settings. However, since our earlier extension of the Barndt and Johnson research led to findings that appeared somewhat paradoxical, it was decided to attempt an exact replication of their study, using adolescent subjects institutionalized for delinquency, rather than younger children in a residential psychiatric treatment center.

METHOD

Subjects. There were two groups of subjects in this investigation. One group consisted of 24 boys who were institutionalized in the Rhode Island Training School for boys, and the other group consisted of 30 girls institutionalized in the Rhode Island Training School for girls. The mean age in both groups was between 15 and 16 years, and all of these adolescents had been sentenced by the court for delinquent activities.

Procedure. The procedure followed was previously employed and described by Barndt and Johnson (1955). The subject was given the following instructions:

> I want to see what kind of a story you can tell. I'll start a story and then let you finish it any way you want to. You can make it any kind of a story you wish. Do you understand? I'll start it now.

The story began as follows:

> About three o'clock one bright sunny afternoon in May, two boys [or two girls in the procedure with the female subjects] were walking along near the edge of town . . .

After reading the beginning of this story to the subject, the experimenter said,

> Now you start there and finish the story any way you want to.

The experimenter then recorded the subject's story verbatim. After finishing the story, the subject was asked, "What time was this?" or "How long did the story take to happen?" The scoring system described by Barndt and Johnson was used. The scoring categories were as follows: 1—under 1 hour; 2—1 hour to under 5 hours; 3—5 hours to under 12 hours; 4—12 hours to under 1 week; 5—1 week to under 3 months; 6—3 months or more. The amount of time covered in the stories could be scored very accurately, and agreement between raters was unity.

RESULTS AND DISCUSSION

The results obtained in this study are shown in Table 1. The findings indicate identical mean time orientation scores in the male and female groups of adolescent delinquents. Table 1 also contains findings obtained in independent studies published previously. It is interesting to note that the time orientation scores of the present groups of adolescent delinquents are very similar to scores obtained by Davids and Parenti (1958) in a study of 11-year-old boys who were institutionalized for severe emotional disturbances and 11-year-old normal boys who were attending a public school. Moreover, the time orientation

[1] This research was supported by Small Grant M-4327 (A) from the National Institute of Mental Health, United States Public Health Service. We wish to express our appreciation to Anthony Travisono, who was Superintendent of the Rhode Island Training School for boys, and Donald D. MacDougald, who is Superintendent of the Rhode Island Training School for girls, for facilitating this study.

TABLE 1

COMPARISON OF PRESENT DELINQUENT GROUPS AND GROUPS STUDIED PREVIOUSLY

Measure	Groups					
	Present delinquent girls ($N = 30$)	Present delinquent boys ($N = 24$)	D-P disturbed boys ($N = 30$)	D-P normal boys ($N = 22$)	B-J delinquent boys ($N = 26$)	B-J normal boys ($N = 26$)
Mean story score	2.7	2.7	2.9	2.9	2.8	3.6
Variance	2.4	1.7	2.7	1.4	2.0	1.7
Mean age	15-5	15-2	10-10	10-10	16-8	16-9

Note.—D-P refer to findings from Davids and Parenti (1958). B-J refer to findings from Barndt and Johnson (1955).

scores of the present groups of delinquents do not differ significantly from the scores obtained in the Barndt and Johnson (1955) study of institutionalized adolescent delinquents. Not only are the mean scores very similar, but according to the F test for homogeneity of variance, there are no significant differences among the variance scores. It is noteworthy that the same assessment procedure utilized several years ago with delinquents in the Midwest yields almost identical findings from delinquents in New England.

Comparison of the present groups of delinquents with Barndt and Johnson's (1955) group of normal adolescent boys showed significant differences in keeping with their findings. Our group of delinquent boys was found to be significantly more present oriented than their normal boys ($t = 2.43$, $p = .05$). The difference between our delinquent girls and their normal boys was even more significant ($t = 6.04$, $p = .001$). Thus, on the basis of these studies, it appears that normal adolescents tend to be more future oriented than do male and female adolescents who engage in serious delinquent activities. It should be emphasized, however, that without a more comprehensive sampling design, these results should be viewed as suggestive rather than definitive.

Our previous findings revealed no significant differences in time orientation in normal and emotionally disturbed 11-year-old children. In fact, these earlier findings suggested that present orientation was correlated with stable friendships and future orientation was correlated with personality characteristics of pessimism and resentment. It was speculated that for children in the latency stage it may be "normal" to be concerned mainly about the present and that concern for the future might be an index of personal dissatisfaction and social instability. However, adequate understanding of these perplexing matters awaits an exhaustive developmental study of normal and maladjusted children at numerous stages of maturation and development. It may well be that what is "normal" or socially desirable at one stage of ego development is pathological at some earlier or later stage in the developmental process.

SUMMARY

This paper represents an attempt to relate findings obtained from three independent studies. Using a story completion technique to assess time orientation in imaginal processes, no difference was found between adolescent delinquent boys and adolescent delinquent girls. The measures of time orientation obtained in this study were practically identical with findings obtained in a previous study of adolescent delinquent boys. Moreover, time orientation revealed in the fantasies of these three groups of delinquents was very similar to time orientation evidenced by much younger emotionally disturbed boys and normal boys. Although there were no significant differences among these five groups, all of them were found to be significantly more present oriented than were a group of adolescent normal boys. Thus, the major finding was a significant difference between time orientation in delinquent and normal children.

REFERENCES

BARNDT, R. J., & JOHNSON, D. M. Time orientation in delinquents. *J. abnorm. soc. Psychol.*, 1955, **51**, 343–345.

DAVIDS, A., & PARENTI, ANITA N. Time orientation and interpersonal relations of emotionally disturbed and normal children. *J. abnorm. soc. Psychol.*, 1958, **57**, 299–305.

WALLACE, M., & RABIN, A. I. Temporal experience. *Psychol. Bull.*, 1960, **57**, 213–236.

(Received March 1, 1961)

Announcing---

the Addition of
300 Articles from
Psychology to

The
Bobbs-Merrill
REPRINT SERIES
IN THE
SOCIAL SCIENCES

The outstanding articles from leading academic journals were selected by a committee of Psychologists—Elliot Aronson, University of Minnesota; Gerald Blum, University of Michigan; Albert Hastorf, Stanford University; and Thomas Landauer, Dartmouth College—for addition to The Bobbs-Merrill Reprint Series in the Social Sciences, which now contains over 1,000 Reprints from Anthropology, Political Science, Psychology and Sociology.

PRICE:

25¢ for articles 32 pp. or less
50¢ for articles 33 pp. to 64 pp.

FORMAT:

Each article is reprinted separately as an 8½" x 11" pamphlet, punched to fit a standard two or three ring binder.

Examine this unique Series before you plan your Fall courses. Write for a complete Catalogue and sample Reprint today to:

COLLEGE DIVISION

THE BOBBS-MERRILL COMPANY, INC.
A SUBSIDIARY OF HOWARD W. SAMS & CO., INC.
TWO EAST FIRST STREET • INDIANAPOLIS 4, INDIANA

Symbol of Good Reading:
SPECTRUM
BOOKS

AN ANATOMY
FOR CONFORMITY

by Edward L. WALKER *and* Roger W. HEYNS, *both at The University of Michigan*

This book is an attempt to clarify and simplify a complex social problem by establishing a more understandable definition of conformity as a change in behavior toward a norm or standard in response to social pressure. This definition is then integrated into a general behavior theory formula: Need, Instrumental Act, and Goal. The variables, which influence the amount of conformity or non-conformity, are then related to this simple framework.

February, 1962 *128 pp.* *(S-24)*
Paperbound: $1.95 *Cloth: $3.95*

HEROES,
VILLAINS, AND FOOLS:
The Changing American Character

by Orrin E. KLAPP, *San Diego State College*

Dr. Klapp's new book examines role-models and stereotypes—Good Guys and Bad Guys—as mirrors of American character and society; the meaning of the "Good Joe"; the compensatory function of the hero; shifting hero-villain identities; alienation and pseudo-integration in contemporary society.

March, 1962 *192 pp.* *(S-31)*
Paperbound: $1.95 *Cloth: $3.95*

For a complete SPECTRUM catalogue, write: DEPT. CAC

Or for approval copies of the above books, write: BOX 903

PRENTICE-HALL, INC.
Englewood Cliffs, New Jersey

"The new and different in tests. . . ."

Could you use objectively obtained estimates of strengths and interrelationships of the id, ego, and superego in your diagnostic work? In your research on personality? Are you just plain curious about how these basic concepts of psychoanalytic theory might be quantified?

The IES TEST

utilizes four independent subtests to measure important aspects of id, ego, and superego functioning. Materials are easy to handle. Tasks are disarming and are intrinsically interesting to almost anyone. Evaluation takes as much time, thought, and sophistication as for any other test giving important basic information about the personality.

•

Obtain information and/or order (complete set $28.50) from

PSYCHOLOGICAL TEST SPECIALISTS
Box 1441
Missoula, Montana

"The new and different in tests. . . ."